JOURNAL OF TELECOMMUNICATIONS
AND INFORMATION TECHNOLOGY

Preface

This issue of the *Journal of Telecommunication and Information Technology* is devoted mostly to applications of multiple criteria decision making, data mining, knowledge acquisition, and other advanced information technologies to telecommunications and other network services, but includes also a paper on strategies of telecommunication technology migration, one on measuring highly doped optical fibers and one on e-learning.

The first paper by Krzysztof Bareja and Włodzimierz Ogryczak from the Warsaw University of Technology presents an outline of a decision support framework based on the concept of reference distributions. This concept is particularly suitable for decision problems with a large number of criteria that have the same character, such as quality of service perceived by each individual user of a telecommunication network, or any other system which serve many users where quality of service for every individual user defines the criteria; in particular, such problems arise when analyzing facility location problems. Distribution of such criteria for a given design of the system can serve as a kind of criteria trajectory and be compared with a reference trajectory or distribution; diverse techniques of fair multicriteria optimization can be used then to select the best design of the system.

The second paper by Andrzej P. Wierzbicki from the National Institute of Telecommunications in Warsaw, based also on a cooperation with the Japan Advanced Institute of Science and Technology in Nomi, Ishikawa, addresses the issue of objectivity versus subjectivity, stressing that while an absolute objectivity is not attainable for diverse reasons, nevertheless trying to be as objective as possible constitutes a higher value, necessary for hard science and technology. Dangers and errors of the subjectivist reduction of objectivity to power and money attempted by the postmodern sociology of science are discussed. On this basis, the problem of subjective versus objective decision analysis and ranking is considered. While all classical decision theory aims at a rational analysis and support of subjective decisions, there are important application cases, particularly in managerial problems, when the decision maker prefers to avoid specifying her/his preferences and needs decision analysis – e.g., ranking of decision options – that is as objective as possible. An approach to decision support that might be easily adapted for such objective ranking is the reference point methodology, such as used in the first paper; its application is shown on real life examples.

The third paper by Janusz Granat and Andrzej P. Wierzbicki from the National Institute of Telecommunications in Warsaw, based also on cooperation with the Warsaw University of Technology, addresses a similar issue to the first paper, but concentrates on the issue of classification of empirical probability distributions (histograms) irrespective of the issue of fairness; generally, classification of histograms is useful both in management situations and in event detection or event mining. While existing approaches to event detection concentrate on the use of selected moments or other characteristics of empirical probability distributions, the paper is based on the postulate that histograms preserve more of needed information then selected moments of this distribution, thus multiple criteria classification of histograms can be most effective in event detection. This paper uses also the concept of objective classification, addressed in the second paper.

The fourth paper by Piotr Rzepakowski from the Warsaw University of Technology uses conjoint analysis for supporting telecommunication product or service sales. Conjoint analysis is widely used as a marketing research technique to study consumers' product preferences and simulate customer choices. The paper addresses the possibility of using conjoint analysis in telecommunication field, to find optimal services which could be recommended to telecommunication customers. A corresponding decision problem is defined. The conjoint analysis method and its connections with ANOVA as well as regression techniques are presented. Diverse utility functions that represent preferences for voice, SMS, MMS and other net services are formulated and compared. Parameters of the proposed conjoint measures are determined by regression methods running on behavioral data, represented by artificially generated call data records. Finally, users are split in homogenous groups by segmentation techniques applied to net service utilities derived from conjoint analysis. Within those groups statistical analyses are performed to create product recommendations.

The fifth paper by Xiaoning Shi and Stefan Voß from the University of Hamburg, based also on a cooperation with Shanghai Jiao Tong University, describes the development of global logistic services in the form of service networks. Shipping companies and liner shipping alliances have moved from pure transportation companies to logistics service providers during the last two decades. Top liner carriers cooperate with their offspring companies that provide local booking service and third party logistics service, combining the business advantages of tight linkages with liners with the negotiation freedom when dealing with demanding customers by providing diverse extensions of service coverage. The paper applies and combines ideas from the Steiner tree problem and game theory to provide a theoretical framework for network design oriented decision support systems of relevant transportation actors and companies. The relation to decision support for telecommunication system network design is also outlined.

The sixth paper by Sylwester Laskowski from the National Institute of Telecommunications in Warsaw addresses the problem of negotiations on regulated markets, such as telecommunication service market, from a game theoretical perspective. It is assumed that two players have to compete with each other on the retail market and cooperate on the wholesale market. The wholesale market is regulated. The role of the regulator is to support players in negotiations, especially by introducing a recommended solutions when the negotiations were broken off. Two cases are considered. First, a situation when one player makes a retail decision before negotiations on wholesale market, and second, a situation when a retail decision follows the negotiations. It is considered how introducing a recommended solutions influence diverse aspects of the negotiation power of the players.

The seventh paper by Szymon Jaroszewicz from the National Institute of Telecommunications in Warsaw addresses cross-selling models for telecommunication services. Cross-selling is a strategy of selling new products to a customer who has made other purchases earlier. This paper presents an analysis of two approaches to cross-selling in a telecommunications setting. The first approach is based on constructing a Bayesian network representing customer's behavior and using this network to predict which customers are most likely to pick each service offered. This gives not only a cross-selling model, but also allows the analyst to gain insight into the behavior of customers. The method starts with a (possibly empty) network representing users background knowledge. At each iteration in finds patterns whose probabilities in customer data diverge most from what the network predicts. An update of the network by an analyst (thus human-computer interactive) corrects for those discrepancies. The second approach uses a separate classifier model for each service offered. Each model predicts, which customers are most likely to buy a specific product. Each customer is then offered a service the classifier of which gives the highest probability of acceptance. The method does not give any insight into customer behavior but is fully automatic.

The eighth paper by Wojciech Michalski from the National Institute of Telecommunications, Warsaw, addresses a survey of next generation network migration profiles deployed by selected carriers in Europe, North America and Asia. The profiles of selected carriers are used as examples of alternative network migration strategies. The paper describes technical status of carriers' networks when the migration process began, the transformation process and its consequences. Strategies were selected mainly from the point of view of carrier's goals and influence of factors like cost of maintaining the public switched telephone network (PSTN), competition and development of voice over IP (VoIP) and multimedia services market. While each operator develops its own unique network migration path, carrier strategies are categorized in this paper into three main groups.

The ninth paper by Krzysztof Borzycki from the National Institute of Telecommunications, Warsaw, addresses the issue of testing highly doped and photonic crystal optical fibers. The paper discusses optical measurements – related to spectral loss, OTDR and PMD, temperature cycling and mechanical tests such as bending, twist and crush, performed on Yb-doped single mode fibers and small-core photonic crystal fibers (PCF). Several issues related specifically to characterization of such special fibers, like measurement errors and artifacts as well as coupling of test instruments to samples are presented. Of particular importance is reliable and low-loss fusion splicing of special fibers to standard single mode fibers (SMF), since most commercially available fiber test instruments are fitted only with SMF interfaces.

The tenth paper is presented by Alina Stasiecka, Ewa Stemposz, and Andrzej Jodłowski from the Polish-Japanese Institute of Information Technology, Warsaw, Institute of Computer Science, Polish Academy of Sciences, Warsaw, and the Higher School of Entrepreneurship and Social Sciences, Otwock, and concerns e-learning resources versus traditional teaching models. The paper presents a discussion of the structure of e-learning resources and its influence on the resources' quality. The thesis of this paper is that the conformance of e-learning resource structure with structures suggested by traditional teaching model/models has a strong impact on the quality of this e-learning resource. The most popular teaching models are analyzed and a proposal of a metamodel useful for e-learning resource construction is introduced.

Andrzej P. Wierzbicki
Guest Editor

Paper

Reference distribution based decision support platform

Krzysztof Bareja and Włodzimierz Ogryczak

Abstract—There many decision problems where numerous partial achievement functions are considered impartially which makes the distribution of achievements more important than the assignment of several achievements to the specific criteria. Such models are generally related to the evaluation and optimization of various systems which serve many users where quality of service for every individual user defines the criteria. This applies to various technical systems, like to telecommunication ones among others, as well as to social systems. An example arises in location theory, where the clients of a system are entitled to equal treatment according to some community regulations. This paper presents an implementation of decision support framework for such problems. This platform is designed for multiple criteria problems analyzed with the reference distribution approach. Reference distribution approach is an extension of the reference point method.

Keywords*— *multiple criteria optimization, decision support systems, reference point method, reference distribution method.

1. Introduction

In various systems which serve many users there is a need to respect the fairness rules while optimizing the total system efficiency. In such problems, the decisions often concern meeting the users' demands in an impartial way. Thus, we are interested rather in distributions of outcomes than specific outcomes themselves. For instance, having two possible location patterns generating for 3 clients outcome vectors (5,0,5) and (0,1,0), respectively, we would recognize both the location patterns as efficient in terms of outcomes (distance) minimization. Indeed, neither (5,0,5) dominates (0,1,0) nor (0,1,0) dominates (5,0,5). However, the first location pattern generates two outcomes (distances) equal to 5 and one outcome equal to 0, whereas the second pattern generates one outcome equal to 1 and two outcomes equal to 0. Thus, in terms of the distribution of outcomes the second location pattern is clearly better. This applies to the desired system output (amount, quality of services) as well as to the obnoxious outcomes (like risk exposure, pollutions). The so-called minimax solution concept, where the worst individual effect (maximum individual disutility) is minimized, is usually considered as the simplest fair optimization model. The minimax approach is consistent with Rawlsian theory of justice [1], especially when additionally regularized with the lexicographic order. On the other hand, making the decisions to optimize the worst individual disutility may cause a large worsening of the overall (mean) performances. Therefore, several other fair decision schemes are searched and analyzed [2, 3, 4].

In this paper we use an alternative concept of the conditional mean which is a parametric generalization of the worst outcome taking into account the portion of population (demands) affected by the worst effects [5]. Namely, for a specified portion β of population we take into account the entire β portion (quantile) of the worst outcomes and we consider their average as the (worst) conditional β-mean outcome. According to this definition the concept of conditional mean is based on averaging restricted to the portion of the worst outcomes. When parameter β approaches 0, the conditional β-mean tends to the worst outcome. On the other hand, for $\beta = 1$ the corresponding conditional mean becomes the standard mean. We select several conditional means for various levels $\beta_k (k = 1, 2, \dots K)$ to get a multiple criteria model, allowing to generate various fair efficient solutions.

Usually there exist many nondominated achievement vectors and they are incomparable with each other on the basis of the specified set of objective functions. Therefore, usually there exist many efficient solutions and they are different not only in the decision space but also in the criteria space. So, there arises a need for further analysis, or rather decision support, to help the decision maker (DM) in the selection of one solution for implementation. Of course, the original objective functions do not allow one to select any efficient solution as better than any other one. Therefore, this analysis depends usually on additional information about the DM's preferences. The DM, working interactively with a decision support system (DSS), specifies the preferences in terms of some control parameters and the DSS provides the DM with an efficient solution which is the best according to the specified control parameters. For such an analysis, there is no need to identify the entire efficient set prior to the analysis, since contemporary optimization software is powerful enough to be used online for direct computations at each interactive step. Thus the DSS can generate at each interactive step only one solution that meets the current preferences. Such a DSS can be used for the analysis of decision problems with finite as well as infinite efficient sets. In order to allow the DSS to meet various DM's preferences it is important, however, that the control parameters provide the completeness of the control (c.f., [6]), i.e., that by varying the control parameters the DM can identify every nondominated achievement vector.

Good controllability can be achieved with the direct use of the reference point methodology (RPM) introduced by Wierzbicki [7] and later extended leading to efficient implementations with many successful applications [8].

The RPM approach is an interactive technique allowing the DM to specify the requirements in terms of aspiration and reservation levels, i.e., by introducing acceptable and required values for several criteria. Depending on the specified aspiration and reservation levels, a special scalarizing achievement function is built which may be directly interpreted as expressing utility to be maximized. Maximization of the scalarizing achievement function generates an efficient solution to the multiple criteria problem. The solution is accepted by the DM or some modifications of the aspiration and reservation levels are introduced to continue the search for a better solution. The RPM approach provides a complete parameterization of the efficient set to multicriteria optimization. Hence, when applying the ARBDS (aspiration-reservation based decision support) methodology to the conditional mean criteria, one may generate various fair and efficient solutions. Since the our criteria defined as conditional means depends only on distribution of outcomes the RPM approach represent actually a reference distribution technique.

2. Sample model: location problem

Presented methodology will be shown by sample problem construction process. As example, we have decided to use well known maximin location problem (LP) variant. This problem is well described on literature (e.g., [9]). In this type of problems we can consider each distance between demand point and location as single criterion. Note, that for DM values are not distinguishable.

2.1. Basics

Let us consider a well known problem of maximum demand point-facility distance. Subject function can be written as $\min\big(\max(d_{ij}x_{ij})\big)$ and LP problem form:

$$\min : z\,.$$

Subject to:

$$z \geq y_i \qquad \forall_i\,,$$

$$y_i = \sum_j d_{ij}x_{ij} \qquad \forall_i \qquad (*)\,,$$

$$\sum_{j=1}^{J} x_{ij} = 1 \qquad \forall_i \qquad (*)\,,$$

$$\sum_{i=1}^{I} f_i \leq F \qquad \forall_i \qquad (*)\,,$$

$$x_{ij} - f_j \leq 0 \qquad \forall_{ij} \qquad (*)\,,$$

where:

$i \in \{1 \dots I\}$ – index of demand point;

$j \in \{1 \dots J\}$ – index of possible facility location;

f_j – (binary variable): 1 if location j was chosen to place facility; 0 otherwise;

x_{ij} – (binary variable): 1 if location i is being served by facility located at j;

y_i – distance from demand point i to facility serving this point;

d_{ij} – distance between demand point i and location j; may be also understood as cost;

F – maximal number of facilities (simple restriction while we don't know about location costs).

Equations marked with asterisk are defining attainable set for variables and in further models will be noted as $[x],[y],[f] \in Q$.

In such formulations it is possible that a single demand node location can influence final result providing to solutions worse for majority of location points. Such situation is illustrated on Fig. 1.

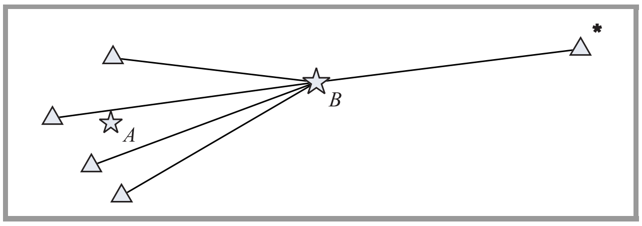


Fig. 1. Unfair domination example.

Chosen facility is considered as worse for almost every demand node. Final result is preferred only for single isolated node in upper right corner of Fig. 1.

2.2. Conditional achievement function

Avoiding "minority dictatorship" described in previous section we can consider value-at-risk like conditional measures. General concept of such models is to minimize the kth worst distance. In example presented above minimizing only 2nd worst distance would result the location A to be preferred.

Simplest formulation of this problem can be written using additional set of binary variables, used for defining set of k worst distances:

$$\min : u_k\,.$$

Subject to:

$$u - y_i \geq -Kz_i\,,$$

$$\sum_{i=1}^{I} z_k \leq k - 1\,,$$

$$[x],[y],[f] \in Q\,,$$

where:

z_i – (binary variable): 1 if distance y_i belongs to $k+1$ maximal distances; 0 otherwise;

K – large value used to switching $u \geq y_i$ inequality;

u – additional variable, equal to kth worst distance in optimal solution;

k – number of value minimized in lexicographically sorted set of distances.

Unfortunately increasing the number of binary variables increases also complexity of presented problem. We can define simpler model based on minimizing total sum of k worst (largest) distances. This approach provides to linear model after transformation to dual problem:

$$\min : \eta .$$

Subject to:

$$\eta = \sum_{i=1}^{I} dd_i + ku \qquad \forall_j ,$$

$$u + dd_i \geq y_i \qquad \forall_i ,$$

$$dd_i \geq 0 \qquad \forall_i ,$$

$$[x], [y], [f] \in Q ,$$

where:

dd_i – downside semideviation between maximal distance and distance node-facility for node i;

η – sum of k worst values in optimal solution.

2.3. Multiple criteria model

By extending concept presented in previous section it is possible to split whole spectrum of distances and minimize every part separately. Granularity of distance distribution description depends on user preferences and in ultimate precision each distance can be modified separately. However, such precision is usually unnecessary. It can also provide to too big model complexity. For example, only number of, i.e., "big", "medium" and "small" values controlling is sufficient.

As result we can consider a model which is "multiplication" of the model described in previous section. General achievement function of this model is not defined yet, but is must be function of conditional measures vector – separate for each interval defined by user:

$$\min : \{\eta_k\} .$$

Subject to:

$$\eta_k = \sum_{i=1}^{I} dd_{ik} + I(k)u_k \qquad \forall_k ,$$

$$y_i = \sum_j d_{ij} x_{ij} \qquad \forall_i ,$$

$$u_k + dd_i \geq y_i \qquad \forall_{i,k} ,$$

$$dd_i \geq 0 \qquad \forall_i ,$$

$$[x], [y], [f] \in Q ,$$

where:

k – the index of interval in lexicographically sorted set of distances;

$I(k)$ – position of demand node related with kth interval.

Each value $I(k)$ is related to respondent partial criteria. In example presented on Fig. 2: $I(0) = 10$ and $I(1) = 20$.

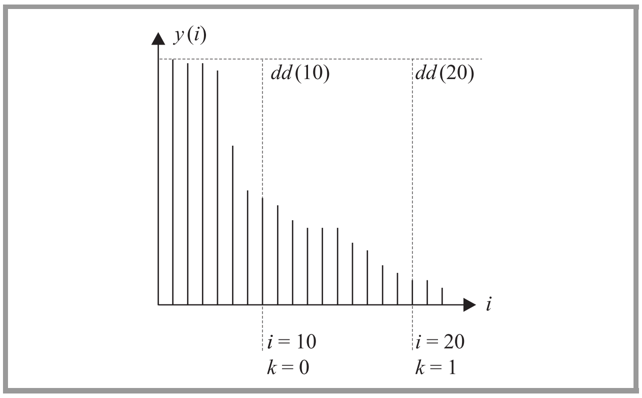


Fig. 2. Example of distances set division.

For the final model two more steps are required: information about user preferences about each interval and some type of scalarization function. While result of given constraint set is a vector of representation, we can treat this vector as set of separate criteria and use the reference point method for this problem.

2.4. Reference point method

For each partial goal we can model preferences using two parameters: level of aspiration (achievement that is fully satisfying decision maker, without need to optimize given criteria) and reservation (minimal). Hence, we can use piecewise linear achievement function, with shape presented in Fig. 3.

Such function can be written as

$$\varphi_k = \max \left[f_1(\eta_k); f_2(\eta_k); f_3(\eta_k) \right]$$

and in terms of LP constraints:

$$\varphi_k \geq a\eta_k - aA_k \qquad \forall_k ,$$

$$\varphi_k \geq \left(\frac{1}{R_k^+ - A_k} \right) \eta_k + \left(\frac{A_k}{A_k - R_k^+} \right) \qquad \forall_k ,$$

$$\varphi_k \geq b\eta_k - bR_k^+ + 1 \qquad \forall_k ,$$

where:

A_k – aspiration value for interval k;

R_k^+ – reservation value for interval k;

a – parameter related to reward for achievement lower than decision maker's aspiration point (value between 0 and 1, arbitrarily defined);

b – parameter related to punishment for achievement bigger than decision maker's reservation point (value greater than 1, arbitrarily defined).

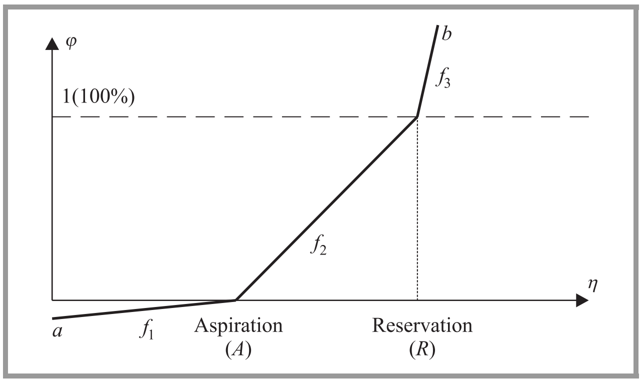


Fig. 3. Piecewise achievement function for single criteria.

Typical generic scalarizing achievement function takes the following form:

$$\min : \max\left(\varphi_k\right) + \varepsilon \sum_{k=1}^{K} \varphi_k$$

and will be used in implemented model.

2.5. User preferences definition

Unfortunately, aspiration and reservation values used in given equations are corresponding to cumulative value of k greatest distances. Thus, aspiration and reservation values should be understood in terms of "total sum of distances larger than kth one". This approach is not intuitive solution for decision maker. We've decided allow user to define preferences in terms of minimal/maximal *average* value in quantile interval (α_k, ρ_k). Proposed translation function is basing on defining increments between beginning and end of described interval. This can be formulated by recursion:

$$A_k = A_{k-1} + \alpha_k\big(I(k) - I(k-1)\big)\,,$$

$$R_k = R_{k-1} + \rho_k\big(I(k) - I(k-1)\big)\,,$$

with initial step:

$$A_0 = \alpha_0 I(0)\,,$$

$$R_0 = \rho_0 I(0)\,.$$

This approach provides to cross dependency between each criterion, and may provide to instability of compromise solution searching process. However, testing the stability is one of implementation goals and this formula may change during test phase.

2.6. Final model formulation

Due to given information final form of LP model is:

$$\min : z + \varepsilon \sum_{k \in (1 \ldots I)} \varphi_k\,.$$

Subject to:

$$z \geq \varphi_k \qquad \forall_k, \qquad (1)$$

$$\varphi_k \geq a\eta_k - aA_k \qquad \forall_k, \qquad (2)$$

$$\varphi_k \geq \left(\frac{1}{R_k^+ - A_k}\right)\eta_k + \left(\frac{A_k}{A_k - R_k^+}\right) \qquad \forall_k, \qquad (3)$$

$$\varphi_k \geq b\eta_k - bR_k^+ + 1 \qquad \forall_k, \qquad (4)$$

$$\eta_k = \sum_{i=1}^{I} dd_{ik} + ku_k \qquad \forall_k, \qquad (5)$$

$$y_i = \sum_j d_{ij}x_{ij} \qquad \forall_i, \qquad (6)$$

$$u_k + dd_i \geq y_i \qquad \forall_{i,k}, \qquad (7)$$

$$dd_i \geq 0 \qquad \forall_i, \qquad (8)$$

$$y_i = \sum_j d_{ij}x_{ij} \qquad \forall_i, \qquad (9)$$

$$\sum_{j=1}^{J} x_{ij} = 1 \qquad \forall_i, \qquad (10)$$

$$\sum_{i=1}^{I} f_i \leq F \qquad \forall_i, \qquad (11)$$

$$x_{ij} - f_i \leq 0 \qquad \forall_{ij}, \qquad (12)$$

where constraints (1)–(4) are responsible for reference point method criteria scalarization, (5)–(8) for definition of criteria η_k and (9)–(12) – for setting up attainable set for variables.

2.7. Model construction conclusions and reservations

In this section we have shown methodology of constructing LP models that allows optimize described class of problems by specialized solver software. However, there are three more issues that require additional discussion.

- Reservation and aspiration values used for piecewise achievement function definition for any particular criterion (c.f. Subsection 2.4) can take any values. Especially there is no condition that aspiration nor reservation values must summarize to one. This fact does not allow to call presented preferences model as distribution.

- As described in Subsection 2.3 single minimized achievement function for interval k represents not distances from defined interval, but from greatest distance to distance represented by kth interval extreme. As result of this fact, implemented preferences model is only approximation. Precision of this approximation increases with number of defined intervals. Preferences model can be assumed as precise for number of intervals equal to number of demand points.

- Construction of partial achievement function definition suggests possible problems with model controllability for the highest distances intervals as affected by every partial achievement function.

3. Refbeans application: multiple criteria decision support framework

Refbeans is an application implemented to presented preferences modeling approach tests. Main purposes of its creation are:

- Methodology demonstration.
- Methodology behavior verification in real life problems, including:
  - attempt to find user friendly optimization process implementation,
  - attempt to find most user friendly way of preferences definition,
  - model controllability verification.
- Creation of easily extensible platform that can serve as a base for future model implementations.

3.1. Project overview

We have decided to implement application as standalone desktop application. Assumed LP form of solved problems allowed to separate frontend part of application, responsible for user interface and data management and library responsible for optimization.
Java language was chosen for base application implementation – mostly because its popularity, proven performance. Big advantage of this choice is existence of two powerful libraries for rich client application implementation: Eclipse RCP and Netbeans (NB) platform. Both libraries has grown from Java IDE's (integrated development environments) and provides wide set of functions related with multi-document edition and management. Our final choice was to make Netbeans as base for our application. It is using native for Java Swing library, what allows to use wide set of existing graphical user interface (GUI) components. JFreeChart library used for chart generation is good example. As persistence layer file storage with extensive use of Java serialization mechanism was used.
Refbeans platform does not provide any LP model implementation. Each LP model must be implemented in as separate Netbeans plugin project, which should depend on framework specific service provider interfaces (SPI's). However at current moment only sample location problem is implemented. Basic platform can be understood as framework which provides common functionality for separate problem implementation. This includes:

1) support for distribution-based preferences definition in three options: graphical input for distribution and cumulative distribution and table form distribution input;
2) support for management of multiple project types, including project data persistence;
3) support for solver communication, including object – MPS translations and data analysis;
4) calculation history management.

As mentioned before, optimization process is performed by external solver, which has to be installed separately. While communication between application and solver is being performed by the MPS file format virtually any solver can be used. However, we recommend GNU linear programming kit (GLPK) [10] solver as first choice while all calculations and tests were performed with this library.

3.2. Application user interface

Application is implemented as stand-alone program. Depending on distribution it can be launched as local program or via Java WebStart technology [11]. User interface is basing on concepts provided by Netbeans platform and similar to other document-oriented applications. Central part of the main window contains document area (Fig. 4). Each opened project is represented by single tab. Detailed single project tab layout, including sub-tabs, charts and tables depends on project type implementation. At system start user is greeted by simple welcome message.

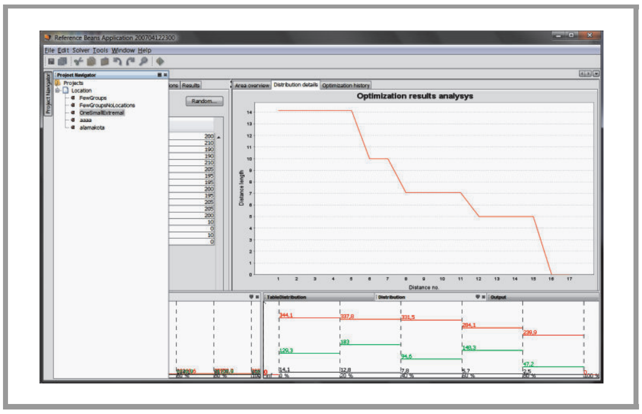


Fig. 4. Application GUI overview.

Project navigator view (minimized, accessible by button on the left window side in default layout) allows to manage prepared projects. Standard CRUD (create, read, update and delete) operations are implemented. Additionally it is possible to export/import project files, what allows to transfer calculation result between computers.
Additional views are coordinated with currently active project. Most important three are being used to specify user criteria. User has possibility to define preferences by distribution chart adjustment, cumulative distribution chart and table – style detailed preferences specification. This allows to test "user friendliness" of different preferences

specification process which is one of application implementation goals. Additionally output view is implemented for providing detailed information about optimization process. This includes mps file view, process performance data view and others. General assumption made on user interface (UI) and code design was, that exact definition of partial achievement parameters is defined by model. This includes definition of intervals extremes. In other words choice if user is minimizing separately worst 10% criteria and rest of 90% separately of worst 90% and 10% best is implemented specific by project type implementation.

Optimization process is being started from main application toolbar. Solver connection must be configured before. This action is available through menu bar.

3.3. Location model implementation

Using common for rich client extendable architecture, application can easily support many types of projects without core source code modifications. Additional modules are implemented as plugins, handled by internal NB platform libraries. First model implemented within presented platform is location problem variant described in Section 2. It operated on two dimensional area with geometrical distance definition. In this implementation user is allowed to define set of demand points and possible facility location by specifying their coordinates (Fig. 5).

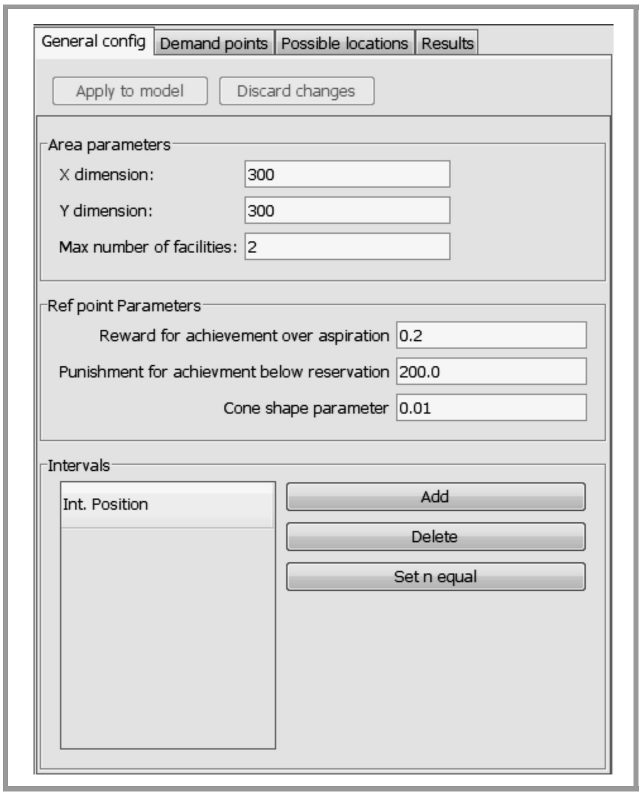


Fig. 5. Location problem parameter definition panel.

Detailed set of parameters describing single problem defined in project implementation includes:

- Area description parameters:
  - area size (width and length),
  - maximum numbers of facilities, that can be established on whole area,
  - list of demand points (X and Y coordinates for each),
  - list of possible facility locations (X and Y coordinates for each).
- Reference point methodology related parameters:
  - reward for partial achievement below aspiration level,
  - "punishment" for exceeding reservation level by partial achievement variable,
  - cone shape related parameter.
- User preferences related parameters:
  - definition of reference distribution discretization points (in terms of percent of total number of criteria/demand points),
  - aspiration and reservation value for each interval.

Detailed data presentation was implemented. User can visualize defined sets of demand and location points on specialized diagram. Although detailed charts of partial achievement goals distribution and optimization history are available.

4. Numerical experiments

4.1. Test data sets

For experimental reasons three artificial data sets were created. First two of them intend to imitate simple location situations for presentation model decisions. Example three and four includes randomly generated larger sets of data for testing model controllability and performance.

Single extreme group case (Fig. 6). This case is created for simple presentation of model ability to act as simple conditional risk models. It was run with one max allowed number of facility location.

Multiple groups of high demand case (Fig. 7). This dataset defines number of separated group of demand points. Available locations are randomly spread all over defined area (with equal probability in every point of area) to make illusion of free choice of location placement.

Volatile demand density model case (Fig. 8). This dataset is intend to check if model is able to recognize areas of high demand points density in noisy environment. Dataset was created by adding large amount of demand points randomly spread all over the area.

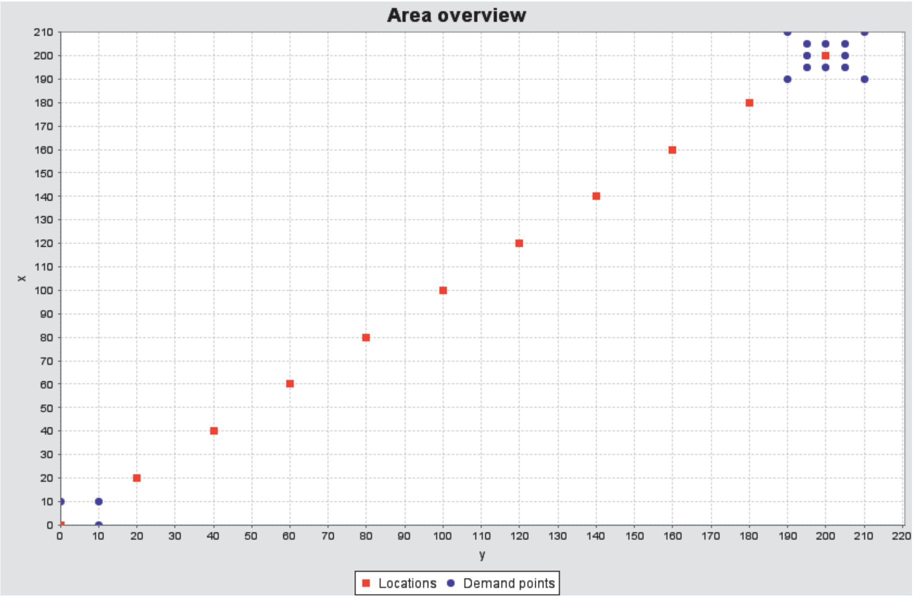


Fig. 6. Single extreme group test case.

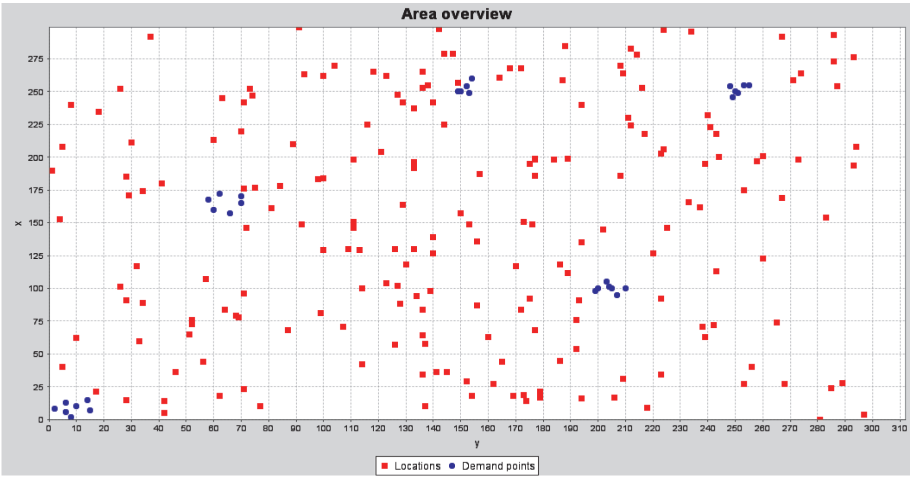


Fig. 7. Multiple groups of high demand test case.

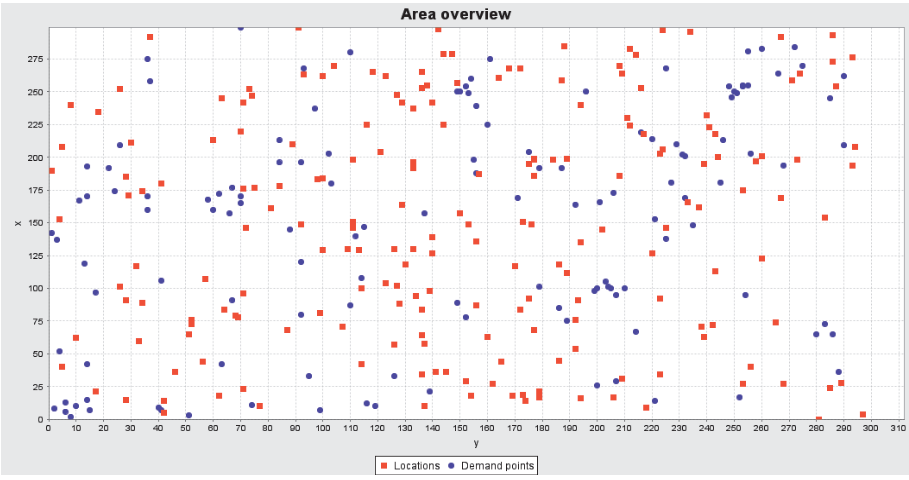


Fig. 8. Volatile demand density model case.

4.2. Model behavior

All presented datasets was successfully optimized, according to authors' assumptions. Quite spectacular was first described set. During experiments we were able to choose continuously any location between center of all possible locations (minimizing the maximal distance) and center of bigger demand group (for preferences specification that completely discards first four largest distances).

Second set was optimized correctly without any need of preferences adjustment. However, this happen because of additional knowledge about model provided in criteria. Number of possible locations was set to 5 – equal to real number of demand point subgroups in dataset. Increasing possible number of location that could be established placed additional location close to one of already selected. Surprisingly we were not able to specify preferences for predictable solution after decreasing number of possible locations. For model it was natural to place locations in one of demand point high density group, leaving one of group being served by "other town" rather than placing location in the middle between two groups. Third set was also optimized correctly, but it required quite many optimization process iterations.

4.3. Model controllability and preferences specification process

General model behavior was positive. Decision maker was able to specify criteria according to presumptions. However, presented implementation leaves big area to improvements. Performed calculations has shown, that model can be too stable and difficult for controlling. It was common to observe no response for user preferences change. This can be explained by discrete nature of presented model. One of possible solution for such behavior is extending aspiration/reservation value range. Implemented version is restricted to distance from zero to maximum possible achievement realization (maximum distance possible on given area in this case). Releasing this constraint will allow to specify preferences in wider range. Thus user should be able to set some intervals as more important by de facto increasing their weight. Models based on continuous domain should be not affected by this symptom.

Model controllability increases with number of possible locations defined. This can be well observed on first data set. For basic data set with 17 possible locations there were practically two local minimums available. This behavior seems to be quite natural – model nature is discrete. Hence each of available solution can be understood as local minimum. Increasing number of possible locations makes distance between each available solution lower and gives DM more flexibility.

Another problem, that provides to (subjective) overcontrollability is lack of information about active restriction. Due to used reference point methodology problem linearization in most situations only single criterion is active and allows to slightly result adjustment. Clear information about active restrictions should be very useful for user and will be probably implemented in future version of application. In current version DM can only find this restriction by experiments.

Another area that require improveents is partial achievement specification. In current implementation it is part of project type implementation and provides not very user friendly interface. This is result of authors' underestimation of distribution intervals manipulation importance.

4.4. Preferences specification model

As described in Subsection 3.1 three ways of preferences definition was implemented. User preferences can be specified in two graphical manners (by specifying distribution or cumulated distribution) and by specification of preferences model distribution values in table organized view (Fig. 9).

Fig. 9. Three available preferences specification interfaces.

Cumulative distribution bases GUI was found difficult to use. This category of achievement definition in location variant problem was not intuitive. In most cases graphical distribution input view was sufficient to found optimal solution. For such preferences model detailed information about aspiration/reservation values importance was not big. Differently from typical reference point models most important information in presented model are relative differences between separate intervals preference values. Of course this conclusion is subjective and cannot be arbitrary while program was tested on small amount of users. Most comfortable way of criteria specifying also varies for different problem types. However implemented preferences specification interface still lacks some data important for DM during optimization process. This includes feedback data from single optimization iteration mentioned in previous paragraph.

Preferences definition interface is generally the most crucial element for proper optimization process. Implemented approach uses different data partial achievement specified by cumulative distribution. This is not intuitive data for user working with implemented location problem. Translation function defined in Subsection 2.5 allows user to specify preferences in terms of distribution but it provides to only rough preferences estimation. Relation between element single achievement and preferences and achievements defined for other intervals is not clearly presented to user. So adjusting preferences with cumulative distribution view allows for more precise result manipulation, despite meaning of the specified preference parameters may be not fully understood by the user.

4.5. Calculation performance

Number of calculations was performed for verifying model size influence on calculation speed. All calculations were performed on Intel Core 2 Duo CPU T7500 model working with 2.20 GHz clock. Java VM used to run application is Sun HotSpot Client VM build 1.6.0_02 working on Windows Vista. Solver used for optimization is pre compiled GLPK solver version 4.9. Performance results can be affected by algorithm implementation which does not take advantage from dual core processor (in both most significant steps: problem file generation and GLPK solver allocation does never use more than one core).

With assumption of about maximum single iteration optimization time two minutes largest solvable models contained about 400 possible location definitions, with 31 demand points and number of total available locations – 5. For similar problem with 100 specified possible locations total computer time exceeded 15 minutes what is too large number for single interactive process iteration.

This paragraph has shown that typical PC available on market for small office and home office (SOHO) consumer is able to perform calculations for middle sized problems without big delay for decision maker. Probably using commercial solver would noticeably improve calculation performance. According to code profiler analysis total time of problem solving can be decreased – major part of time is take by problem MPS file creation. This part of code is custom made and leaves area for code optimization.

5. Conclusions

In this paper we have described class of multicriteria problems for which preferences can be modeled in terms of distribution. Proposed preferences modeling technique is an variation of the reference point methodology and is using similar aspiration/reservation concepts. We have also shown method of construction LP problems basing exemplary location problem.

For model presentation a demonstration application was implemented. Refbeans platform allows us to test presented approach in real-life simulation problem. At current time only single type of problems was implemented. Despite this, platform architecture should allow to implement other types of problems in relatively easy way.

Performed experiments has shown, that proposed preferences modeling technique can be successfully used for optimal solution search. Proposed process is basing on adjusting user preferences in iterations. However, model response for preferences value changes is weak. This subject was discussed in Subsection 4.4. Preferences specification interface also requires improvements. Especially number of information presented to DM as optimization result should be increased.

Both model and application implementation has some flaws. Some of them was presented in previous sections. However, current status allows to improve them in future research. Decision about model usability in enterprise environment cannot be made basing only on the current status.

Appendix

Invitation to experiments and contribution

Main purpose of presented application implementation is test and demonstration of presented modeling approach. We would like to encourage users to participate in both experiments and platform development. For running it system must have Java runtime environment installed with minimum version 1.5. System is using external solver for calculations, which you have to download separately. It can be any MPS file type aware solver, but we recommend GLPK solver. All development-stage tests were performed on it. One can download GLPK solver from any GNU software mirror site. One can also download pre-compiled version for Windows.

Application source is also available under free license. Please contact the authors for full source access and support.

Acknowledgments

The research was supported by the Ministry of Science and Information Society Technologies under grant 3T11C 005 27 “Models and Algorithms for Efficient and Fair Resource Allocation in Complex Systems”.

References

[1] J. Rawls, *The Theory of Justice*. Cambridge: Harvard University Press, 1971.

[2] R. Denda, A. Banchs, and W. Effelsberg,“The fairness challenge in computer networks”, in *First COST 263 International Workshop, Quality of Future Internet Services, QofIS 2000, Berlin, Germany, 2000,* J. Crowcroft, J. Roberts, and M. I. Smirnov, Eds. Berlin: Springer, 2000, vol. 1922, pp. 208–220.

[3] M. M. Kostreva and W. Ogryczak, "Linear optimization with multiple equitable criteria", *RAIRO Oper. Res.*, vol. 33, pp. 275–297, 1999.

[4] W. Ogryczak and T. Śliwiński, "On equitable approaches to resource allocation problems: the conditional minimax solution", *J. Telecommun. Inform. Technol.*, no. 3, pp. 40–48, 2002.

[5] W. Ogryczak and M. Zawadzki, "Conditional median: a parametric solution concept for location problems", *Ann. Opns Res.*, vol. 110, pp. 167–181, 2002.

[6] A. P. Wierzbicki, M. Makowski, and J. Wessels, *Model Based Decision Support Methodology with Environmental Applications*. Dordrecht: Kluwer, 2000.

[7] A. P. Wierzbicki, "A mathematical basis for satisficing decision making", *Math. Model.*, vol. 3, pp. 391–405, 1982.

[8] A. Lewandowski and A. P. Wierzbicki, *Aspiration Based Decision Support Systems – Theory, Software and Applications*. Berlin: Springer, 1989.

[9] S. H. Owen and M. S. Daskin, "Strategic facility location: a review", *Eur. J. Opnl Res.*, vol. 111, pp. 423–447, 1998.

[10] A. Makhorin, "GNU linear programming kit", http://www.gnu.org/software/glpk/

[11] Sun Microsystems Inc., http://java.sun.com/products/javawebstart/

[12] M. M. Kostreva, W. Ogryczak, and A. Wierzbicki, "Equitable aggregations and multiple criteria analysis", *Eur. J. Opnl Res.*, vol. 158, pp. 362–367, 2004.

Krzysztof Bareja received his M.Sc. in computer science from the Warsaw University of Technology, Poland, in 2005. Currently he is a Ph.D. student in the Institute of Control and Computation Engineering at the Warsaw University of Technology. His research area focuses on decision support systems implementation, using Java in desktop systems and application design. His commercial experience includes over five years of work including participation in projects for large telecommunication vendor and Polish government clients.

e-mail: kbareja@elka.pw.edu.pl
Institute of Control and Computation Engineering
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland

Włodzimierz Ogryczak is a Professor and the Head of Optimization and Decision Support Division in the Institute of Control and Computation Engineering (ICCE) at the Warsaw University of Technology, Poland. He received both his M.Sc. (1973) and Ph.D. (1983) in mathematics from Warsaw University, and D.Sc. (1997) in computer science from Polish Academy of Sciences. His research interests are focused on models, computer solutions and interdisciplinary applications in the area of optimization and decision making with the main stress on: multiple criteria optimization and decision support, decision making under risk. He has published three books and numerous research articles in international journals.

e-mail: W.Ogryczak@ia.pw.edu.pl
Institute of Control and Computation Engineering
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland

Paper

The problem of objective ranking: foundations, approaches and applications

Andrzej P. Wierzbicki

Abstract—The paper starts with the discussion of the issue of objectivity versus subjectivity, stressing that while an absolute objectivity is not attainable, nevertheless trying to be as objective as possible constitutes a higher value, necessary for hard science and technology. Dangers and errors of the subjectivist reduction of objectivity to power and money attempted by the postmodern sociology of science are discussed. Then we turn to the problem of subjective versus objective decision analysis and ranking. It is shown that while all classical decision theory aims at a rational analysis and support of subjective decisions, there are important application cases, particularly in managerial problems, when the decision maker prefers to avoid specifying her/his preferences and needs decision analysis – e.g., ranking of decision options – that is as objective as possible. An approach to decision support that might be easily adapted for such objective ranking is the reference point methodology; its application is shown on examples. One of these examples is actually not an application of the methodology, but a real life problem that motivated the development of objective ranking. The examples illustrate that objective ranking might be important for management, including also management of telecommunication networks.

Keywords— subjective ranking, objective ranking, reference point approaches, objectivity.

1. Introduction

The words *subjective* and *objective* might be used in a derogatory sense, but we shall use them in their original epistemic sense:

- subjective as resulting from personal cognition or preferences;
- objective as trying to represent outside world without bias and presuppositions.

Thus, we can say that all contemporary decision analysis, aiming at supporting the decision maker in using her/his own preferences for selecting best personal decisions, concentrates actually on computerized, rational support of subjective decisions. But what means *computerized support*? It should include at least two aspects:

- a computerized representation of knowledge (including data, rules, models) about a part of outside reality pertinent for the decision situation, which should be *as objective as possible*;
- a computerized support for combining *the subjective preferences* of an individual decision maker with an objective representation of the pertinent knowledge in selecting the actual decision.

However, there are practical cases (illustrated by examples given later) when the decision maker does not want to specify her/his individual preferences, prefers to obtain suggested decisions – or a ranking of a list of decision options specified as objectively as possible.

We know that full objectivity is impossible. This was shown already by Heisenberg [8], we discuss it in more detail later – and contemporary physics still considers a synthesis of Heisenbergian *indeterminacy* and Einsteinian *relativity* as the most important problem in science. However, technology and engineering cannot develop without trying to be as objective as possible, for example, without submitting technological tools to destructive Popperian falsification tests.

Postmodern social science ridicules Popperian falsificationism and postulates that all our knowledge is subjective, but we shall discuss the errors of postmodern sociology of science later. Here we just conclude that *there is a need of both subjective and objective aspects of knowledge and decisions*.

2. Objectivity versus subjectivity

At the beginning, we must add some philosophical comments on subjectivity and objectivity. The destruction of the industrial era episteme [25, 28] – sometimes called not quite precisely positivism or scientism – started early, e.g., since Heisenberg [8] has shown that not only a measurement depends on a theory and on instruments, but also the very fact of measurement distorts the measured variable.

This was followed by diverse philosophical debates, summarized, e.g., by Quine [20] who has shown that the logical empiricism (neo-positivism) is logically inconsistent itself, that all human knowledge "*is a man-made fabric that impinges on existence only along the edges*". This means that there is no absolute objectivity.

However, this was quite differently interpreted by hard sciences and by technology, which nevertheless tried to remain as objective as possible, and by social sciences which, in some cases, went much further to maintain that all knowledge is subjective – results from a discourse, is constructed,

negotiated, relativist, depends on power and money (see, e.g., [13]).
This has led to a general divergence of *the episteme* – the way of constructing and justifying knowledge, characteristic for a given cultural era (see [4]), we add only that also characteristic for a cultural sphere – of the following three different cultural spheres (see [25]):

– of hard and natural sciences;

– of technology proper (understood as the art of constructing tools);

– of social sciences and humanities.

Even if we (the technologists) respect the different culture of social sciences and humanities, we must protest against extreme epistemic interpretations that become fashionable today. For example, some our colleagues maintain that "*There is no universe, but only a multiverse – and to realize this is liberating*". We propose that they liberate themselves by falsifying their conviction, applying a hard wall test: posit yourself against hard wall, close your eyes, and try to convince yourself that, since there is only a multiverse (and, according to the quantum theory, there is a nonzero probability of penetrating the wall), the wall does not exist. If you cannot convince yourself, then there is no multiverse, because reality apparently has some universal features; if you can convince yourself, run with your head ahead, in order to falsify your conviction.
On the other hand, even if we should try to develop an integrated episteme for the new era of knowledge civilization (see [28]), this new episteme must take into account that absolute objectivity is not attainable, because of the following basic principles.

Multimedia principle: words are just an approximate code to describe much more complex reality, visual and generally preverbal information is much more powerful and relates to intuitive knowledge and reasoning; future records of the intellectual heritage of humanity will have multimedia character, thus stimulating creativity.
This multimedia principle has many implications, but we stress here only the most obvious: if words are just an approximate code, then absolute truth and absolute objectivity are obviously not possible. But we need truth even in elementary social discourse, need objectivity at least in technology. Thus, truth and objectivity are higher values, ideals that we try to attain as closely as possible even if they are not fully attainable.
This is related to another basic principle. The concepts of punctuated evolution from biology, order emerging out of chaos from computational modeling, emergence of software out of hardware, multiple layers of protocols in telecommunications jointly justify the following.

Emergence principle: new concepts and properties of a system emerge with increased level of complexity, and these properties are qualitatively different than and irreducible to the properties of parts of the system.
This principle implies a fundamental conceptual change. Firstly, it shows that the arguments of creationism against evolution – that evolution could not produce irreducible complexity – are ignorant of the obvious fact that the evolution of civilization, much faster than the biological evolution thus easier to observe, has recently produced several examples of the emergence of irreducible complexity, starting with the emergence of software out of hardware. Secondly, even if it might seem that emergence principle logically results from the principle of synergy or holism – that the whole is more than the sum of its parts (see [1, 2]), this is not necessarily a correct interpretation. The principle of synergy or holism does not say that the whole should have essentially different, irreducible properties, than the parts of the system.
However, we can see that higher values, such as truth or objectivity, are also higher level concepts that emerged evolutionary in civilisation evolution. Thus, they are irreducible to lower level concepts, such as power and money.
This is not just a philosophic debate. If scientific objectivity could be reduced to money and power, than managers would try to force us, engineers, to use fraudulent engineering for profit; and, more generally, postmodern sociology of science gives a nice excuse for an unlimited privatization of knowledge.
The argument for privatization of public resources is based on the phenomenon of tragedy of commons (devastation of a degradable resource, if used without limits). However, knowledge is not degradable (see [14]), it increases with use, hence it is more advantageous for a community to keep knowledge public. But there are strong economic forces today interested in an unrestricted privatization of knowledge; and postmodernism provides them with an ideology.
Thus, we should clearly point out the errors of postmodern sociology of science. For example, Latour [13] argues that since the concepts of nature and reality are constructed by us, they cannot be the cause of our knowledge, because an effect cannot be a cause. His argument is logically erroneous, in obvious ignorance about the mechanism of positive feedback that is the basis of the evolution of knowledge. Hence, it is not true that knowledge can be reduced to money and power.
In order to show that the postmodern episteme is not the only possible one, we present here another description of the relation of human knowledge to nature [28]. First, from a technological perspective we do not accept the assumption of postmodern philosophy that "nature" is only a construction of our minds and has only local character. Of course, the word nature refers both to the construction of our minds and to something more – to some persisting, universal (to some degree) aspects of the world surrounding us. People are not alone in the world; in addition to other people, there exists another part of reality, that of nature, although part of this reality has been converted by people to form human-made, mostly techno-

logical systems. There are aspects of reality that are local and multiple, there are aspects that are more or less universal.

Second, the general relation of human knowledge to reality might be described as follows. People, motivated by curiosity and aided by intuition and emotions, observe reality and formulate hypotheses about properties of nature, of other people, of human relations; they also construct tools; together, we call all this knowledge. People test and evaluate the knowledge constructed by them by applying it to reality: perform destructive tests of tools, devise critical empirical tests of theories concerning nature, apply and evaluate theories concerning social and economic relations; in general, we can consider this as a generalized principle of falsification, broader than defined by Karl Popper even in his later works [19].

Such a process can be represented as a general spiral of evolutionary knowledge creation, see Fig. 1. We observe reality (either in nature or in society) and its changes, compare our observations with human heritage in knowledge (the transition *Observation*). Then our intuitive and emotive knowledge helps us to generate new hypotheses (*Enlightenment*) or to create new tools; we apply them to existing reality (*Application*), usually with the goal of achieving some changes, modifications of reality (*Modification*); we observe them again.

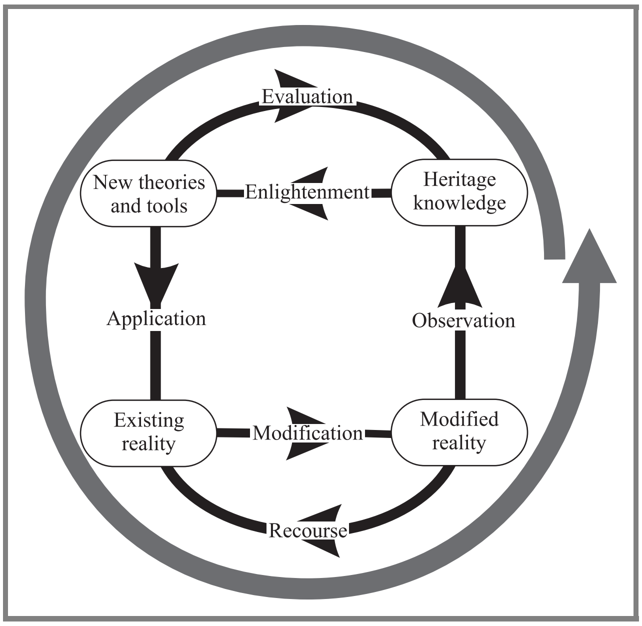


Fig. 1. The general OEAM spiral of evolutionary knowledge creation.

It is important, however, to note that many other transitions enhance this spiral. First is the natural evolution in time: modified reality becomes existing reality through *Recourse*. Second is the evolutionary selection of tested knowledge: most new knowledge might be somehow recorded, but only the positively tested knowledge, resilient to falsification attempts, remains an important part of human heritage (*Evaluation*); this can be interpreted as an objectifying, stabilizing feedback. Naturally, there might be also other transitions between the nodes indicated in the spiral model, but the transitions indicated in Fig. 1 are the most essential ones.

Thus, nature is not only the effect of construction of knowledge by people, nor is it only the cause of knowledge: it is both cause and effect in a positive feedback loop, where more knowledge results in more modifications of nature and more modifications result in more knowledge. As it is typical for positive feedback loops, the overall result is an avalanche-like growth; and this avalanche-like growth, if unchecked by stabilizing negative feedbacks, beside tremendous opportunities creates also diverse dangers, usually not immediately perceived but lurking in the future. Thus, the importance of selecting knowledge that is as objective as possible relates also to the fact that avalanche-like growth creates diverse threats: we must leave to our children best possible knowledge in order to prepare them for dealing with unknown future.

This description of a spiral-like, evolutionary character of knowledge creation is consistent with our technological cognitive horizon, and slightly different than presented in [9] from a position of an economic cognitive horizon. It is an extension of the concept of objective knowledge promoted by Popper [19], but admits relativistic interpretations; it only postulates objectivity as a higher level value, similar to justice: both absolute justice and absolute objectivity might be not attainable, but are important, worth striving for, particularly if we take into account uncertainty about future (see also [21]).

After outlining this philosophic background, we can turn now to the problem of objective versus subjective ranking of decision options. We start, however, with an outline of reference point approaches to the problem of ranking.

3. Reference point approaches: the discrete case

We assume here that the admissible decisions are given by just a list of considered decision options $\boldsymbol{x}_k \in X_0$, where X_0 denotes the set of these options. We assume that we have a decision problem with J criteria, indexed by $j = 1,\dots,J$ (also denoted by $j \in \boldsymbol{J}$), and K decision options called also alternatives, indexed by $k = 1,\dots,K$ or $k = \mathrm{A},\mathrm{B},\dots,\mathrm{H}$ (also denoted by $k \in \boldsymbol{K} = \{1,\dots,K\}$). The corresponding criteria values are denoted by q_{jk}; we assume that all are maximized or converted to maximized variables. The maximal values $\max_{k\in\boldsymbol{K}} q_{jk} = q_j^{up}$ are called upper bounds for criteria and are often equivalent to the components of so called ideal or utopia point $\boldsymbol{q}^{uto} = \boldsymbol{q}^{up} = (q_1^{up},\dots,q_j^{up},\dots,q_J^{up})$ – except for cases when they were established a priori as a measurement scale. The minimal values $\min_{k\in\boldsymbol{K}} q_{jk} = q_j^{lo}$ are called lower bounds and, generally, are not equivalent to the components of so called nadir point $\boldsymbol{q}^{nad} \geq \boldsymbol{q}^{lo} = (q_1^{lo},\dots,q_j^{lo},\dots,q_J^{lo})$; the nadir point $\boldsymbol{q}^{nad}$ is defined similarly as the lower bound point $\boldsymbol{q}^{lo}$, but with minimization restricted to Pareto optimal or efficient or nondominated alternatives (see, e.g., [3]). An alternative $k^* \in K$ is Pareto optimal (Pareto-nondominated or shortly nondominated, also called efficient), if there is no other alternative $k \in \boldsymbol{K}$ that dominates k^*, that is, if we denote $\boldsymbol{q}_k = (q_{1k}, \ldots, q_{jk}, \ldots, q_{Jk})$, there is no $k \in \boldsymbol{K}$ such that $\boldsymbol{q}_k \geq \boldsymbol{q}_{k^*}$, $\boldsymbol{q}_k \neq \boldsymbol{q}_{k^*}$.

While the reference point approach is typically described for the continuous case (with a nonempty interior of X_0, thus an infinite number of options in this set), we shall concentrate here on the discrete case, with a finite number of decision options K, for which case the reference point approach is equally or even particularly suitable. This is because when we consider the outcome set Q_0, that is, the set of criteria vectors $\boldsymbol{q}_k$ corresponding to decision options $\boldsymbol{x}_k$, is in this case obviously not convex, and important elements of Pareto outcome set might be contained in the interior, not on the boundary of the convex cover of this set. Thus, with any other method – particularly with a weighted sum, but also with many nonlinear utility approximations – we run the risk of missing important Pareto points in the discrete case. We do not have this risk when using reference point approaches, because of their full controllability property, not possessed by utility functions nor by weighted sums, see later comments.

The most general specification of preferences of a decision maker contains a selection of decision outcomes chosen as criteria, accompanied by defining a partial order in the space of criteria – simply asking the decision maker which criteria should be maximized and which minimized (or stabilized). Here we consider only the simplest case when all criteria are maximized.

When analyzing a decision problem in the discrete case, we might be interested in:

– finding the best solution (option),

– finding all Pareto-optimal solutions (options),

– ranking all options,

– classifying all options.

Here we shall consider mostly the problem of ranking. There are several versions of methods belonging to the general class of reference point approaches (see [15, 26]). Here we describe a method based on a specification of double reference levels – aspiration level a_j and reservation level r_j – for each criterion. After this specification, the approach uses a nonlinear aggregation of criteria by an achievement function that is performed in two steps.

We first count achievements for each individual criterion or satisfaction with its values by transforming it (monotonically and piece-wise linearly), e.g., in the case of maximized criteria as shown in Eq. (1) below. In a discrete decision problem we can choose these coefficients to have a reasonable interpretation of the values of the partial (or individual) achievement function. Since the range of [0; 10] points is often used for eliciting expert opinions about subjectively evaluated criteria or achievements, we adopted this range in Eq. (1) below for the values of a partial achievement function $\sigma_j(q_j, a_j, r_j)$:

$$\sigma_j(q_j,a_j,r_j) = \begin{cases} \dfrac{\alpha(q_j - q_j^{lo})}{(r_j - q_j^{lo})} & \text{for } q_j^{lo} \leq q_j < r_j, \\ \dfrac{\alpha + (\beta - \alpha)(q_j - r_j)}{(a_j - r_j)} & \text{for } r_j \leq q_j < a_j, \\ \dfrac{\beta + (10 - \beta)(q_j - a_j)}{(q_j^{up} - a_j)} & \text{for } a_j \leq q_j \leq q_j^{up}. \end{cases} \tag{1}$$

The parameters α and β, $0 < \alpha < \beta < 10$, in this case denote correspondingly the values of the partial achievement function for $q_j = r_j$ and $q_j = a_j$. The value $\sigma_{jk} = \sigma_j(q_{jk}, a_j, r_j)$ of this achievement function for a given alternative $k \in \boldsymbol{K}$ signifies the satisfaction level with the criterion value for this alternative. Thus, the above transformation assigns satisfaction levels from 0 to α (say, $\alpha = 3$) for criterion values between q_j^{lo} and r_j, from α to β (say, $\beta = 7$) for criterion values between r_j and a_j, from β to 10 for criterion values between a_j and q_j^{up}.

After this transformation of all criteria values, we might use then the following form of the overall achievement function:

$$\sigma(\boldsymbol{q},\boldsymbol{a},\boldsymbol{r}) = \min_{j \in \boldsymbol{J}} j_i(q_j,a_i,r_j) + \varepsilon/J \sum_{j \in \boldsymbol{J}} \sigma_j(q_j,a_j,r_j), \tag{2}$$

where $\boldsymbol{q} = (q_1, \ldots q_j, \ldots q_J)$ is the vector of criteria values, $\boldsymbol{a} = (a_1, \ldots a_j, \ldots a_J)$ and $\boldsymbol{r} = (r_1, \ldots r_j, \ldots r_J)$ are the vectors of aspiration and reservation levels, while $\varepsilon > 0$ is a small regularizing coefficient. The achievement values $\sigma_k = \sigma(\boldsymbol{q}_k, \boldsymbol{a}, \boldsymbol{r})$ for all $k \in \boldsymbol{K}$ can be used either to select the best alternative, or to order the options in an overall ranking list or classification list, starting with the highest achievement value.

The properties of such functions are, also for the discrete case:

– partial order approximation: the level sets of such functions approximate closely the positive cone defining the partial order (see [24]);

– full controllability: given any point $\boldsymbol{q^*}$ in criteria space that is (properly, with a priori bounded trade-off coefficients[1]) Pareto-nondominated and corresponds to some decision option, we can always

[1] By a properly Pareto-nondominated option with a priori bounded trade-off coefficients, called also an ε-properly Pareto-nondominated alternative, we understand a Pareto-nondominated alterbnative with trade-off coefficients bounded by a given large number, e.g., the number $1 + 1/\varepsilon$ [26]. The property that any ε-properly Pareto-nondominated alternative can be selected as the best by maximizing an achievement function is called the controllability property and is much stronger than the efficiency property (that any maximum of a function, which is strictly monotone with respect to the partial order, is Pareto-nondominated). The controllability property is possessed by functions such as (2) that are not only strictly monotone with respect to the partial order, but also have level sets approximating the positive cone that defines the partial order. This property does not depend on convexity assumptions [24].

choose such reference levels – in fact, it suffices to set aspiration levels equal to the components of $\boldsymbol{q}^*$ – that the maximum of the achievement function (2) is attained precisely at this point;

– dependence of implied weighting coefficients on the currently specified reference points (see [26]);

– possibility of using $\varepsilon = 0$ with nucleolar minimax approach [12, 16]: consider first the minimal, worst individual criterion achievement σ_j computed as above with $\varepsilon = 0$; if two options k_1 and k_2 (or more of them) have the same achievement value, we order them according to the second worst individual criterion achievement, and so on.

4. The issue of objective ranking

The ranking of discrete options is a classical problem of multi-attribute decision analysis; however, all classical approaches – whether of Keeney and Raiffa [11], or of Saaty [22], or of Keeney [10] – concentrate on *subjective ranking*. By this we do not mean intuitive subjective ranking, which can be done by any experienced decision maker based on her/his intuition, but rational subjective ranking, based on the data relevant for the decision situation – however, using an approximation of personal preferences in aggregating multiple criteria.

And therein is the catch: in many practical situations, if the decision maker wants to have a computerized decision support and rational ranking, she/he does not want to use personal preferences, prefers to have some objective ranking. This is often because the decision is not only a personal one, but affects many people – and it is usually very difficult to achieve an intersubjective rational ranking, accounting for personal preferences of all people involved. This obvious fact is best illustrated by the following example.

Suppose an international corporation consists of six divisions A–F. Suppose these units are characterized by diverse data items, such as name, location, number of employees, etc. However, suppose that the Chief Executive Officer (CEO) of this corporation is really interested in ranking or classification of these divisions taking into account the following attributes used as criteria: 1) profit (in % of revenue), 2) market share (m.share, in % of supplying a specific market sector), 3) internal collaboration (i.trade, in % of revenue coming from supplying other divisions of the corporation), and 4) local social image (l.s.i., meaning public relations in the society where it is located, evaluated on a scale 0–100 points). All these criteria are maximized, improve when increased. An example of data of this type is shown in Table 1 (with data distorted for privacy reasons).

Suppose the CEO of this corporation hires a consulting company and asks for an objective ranking of these six divisions. The approach that can be easily adapted for rational objective ranking is reference point approach – because reference levels needed in this approach can be either defined subjectively by the decision maker, or established objectively statistically from the given data set. We can use this approach not only for objective ranking, but also for objective classification, using methods as indicated above with objectively defined reference points.

In the next section, we shall show below how to apply this approach for the simple example given in Table 1. We denoted by q_{jk} the value of a criterion q_j for the decision option $k \in \boldsymbol{K}$, and the achievement values $\sigma_k = \sigma(\boldsymbol{q}_k, \boldsymbol{a}, \boldsymbol{r})$ for all $k \in \boldsymbol{K}$ can be used to order the options in an overall ranking list, starting with the highest achievement value. Now, the question is: how to define aspiration levels $\boldsymbol{a}$ and reservation levels $\boldsymbol{r}$ in order to obtain rational objective ranking? Several ways were listed in [5]: neutral, statistical, voting; we shall concentrate here on statistical determination.

A statistical determination of reference levels concerns values m_j that would be used as basic reference levels, an upward modification of these values to obtain aspiration levels a_j, and a downward modification of these values to obtain reservation levels r_j; these might be defined as follows:

$$m_j = \sum_{k \in \boldsymbol{K}} \boldsymbol{q}_{jk}/K; \; r_j = 0.5(q_j^{lo} + m_j); \; a_j = 0.5(q_j^{up} + m_j), \; \forall j \in \boldsymbol{J}, \tag{3}$$

where K denotes the number of alternative options, thus m_j are just average values of criteria in the set of all alternative options; aspiration and reservation levels – just averages of these averages and the lower and upper bounds, respectively.

However, there are no essential reasons why we should limit the averaging to the set of alternative options ranked; we could use as well a larger set of data in order to define more adequate (say, historically meaningful) averages, or a smaller set – e.g., only the Pareto optimal options – in order to define more demanding averages and aspirations. For very large data sets, we can use, e.g., evolutionary (EMO) algorithms for an approximation of the Pareto set.

Variants of objective ranking. For the data from Table 1, we can thus present two variants of objective ranking:

– A: based on averages of data from this table,

– B: based on averages from Pareto optimal options.

See next Tables 2 and 3. Note that the more demanding ranking B displays a rank reversal: the divisions C and E, occupying positions 2 and 3 in ranking A, exchange their places in ranking B. This is, however, a natural phenomenon: average aspirations favour standard though good solutions, truly interesting solutions result from demanding aspirations (however, this rank reversal might disappear, if we use different values of the parameter ε).

Note also that the rank reversal also disappears if, instead of ranking, we classify the divisions into three classes:

– I: very good,

– II: good,

– III: wanting.

Table 1
Data for an example on international business management (* denotes Pareto options)

Division	Name	Location	Employees	Profit – q_1 [%]	M.share – q_2 [%]	I.trade – q_3 [%]	L.s.i. – q_4
A	Alpha	USA	250	11	8	10	40
B*	Beta	Brasilia	750	23	40	34	60
C*	Gamma	China	450	16	50	45	70
D*	Delta	Dubai	150	35	20	20	44
E*	Epsilon	C. Europe	350	18	30	20	80
F	Fi	France	220	12	8	9	30

Table 2
Example of objective ranking for data from Table 1, based on averages of all options

Ranking *A* Division	σ_1	σ_2	σ_3	σ_4	σ	Rank	Class
A	0.00	0.00	0.37	2.50	0.29	5	III
B	5.63	7.50	7.00	5.88	8.23	1	I
C	3.30	10.0	10.0	7.62	6.39	2	II
D	10.0	3.57	3.89	3.32	5.40	4	II
E	3.97	5.48	3.89	10.0	6.30	3	II
F	0.73	0.00	0.00	0.00	0.07	6	III

Table 3
Example of objective ranking for data from Table 1, based on averages of Pareto-nondominated options

Ranking *B* Division	σ_1	σ_2	σ_3	σ_4	σ	Rank	Class
A	0.00	0.00	0.29	1.80	0.21	5	III
B	5.00	6.61	6.24	5.13	7.30	1	I
C	2.50	10.0	10.0	6.73	5.28	3	II
D	10.0	3.47	3.13	2.51	4.42	4	II
E	3.33	5.04	3.13	10.0	5.43	3	II
F	0.50	0.00	0.00	0.00	0.05	6	III

Both divisions C and E remain in the class II, both for the average and for the more demanding aspirations.

In some management applications, the worst ranked options are the most interesting, because they indicate the need of a corrective action. Objective ranking was originally motivated by an actual application when evaluating scientific creativity conditions in a Japanese research university, Japan Advanced Institute of Science and Technology (JAIST), (see [23]). Actually, it is misleading to call it an application; a real life problem was first solved innovatively, which motivated later the development of theory. This often happens in technology development: technology is not necessarily and not only an application of basic natural science, it often precedes theoretical developments – such as invention of a wheel preceded the concept of a circle, a telescope preceded optics.

The evaluation was based on survey results. The survey included 48 questions with diverse answers and over 140 respondents with diverse characteristics: school affiliation (JAIST consists of three schools), nationality (Japanese or foreign – the latter constitute over 10% of young researchers at JAIST), research position (master students, doctoral students, research associates, etc.). In total, the data base was not very large, but large enough to create computational problems.

The questions were of three types:

– assessment questions, assessing the situation between students and at the university; the most critical questions of this type might be selected as those that correspond to worst responses;

– importance questions, assessing importance of a given subject; the most important questions might be considered as those that correspond to best responses;

– controlling questions, testing the answers to the first two types by indirect questioning revealing responder attitudes or asking for a detailed explanation.

For the first two type of questions, responders were required to tick appropriate responses in the scale $\boldsymbol{vg}$ (very good), $\boldsymbol{g}$ (good), $\boldsymbol{a}$ (average), $\boldsymbol{b}$ (bad), $\boldsymbol{vb}$ (very bad) – sometimes in an inverted scale if the questions were negatively formulated. Answers to all questions of first two types were evaluated on a common scale, as a percentage distribution (histogram) of answers $\boldsymbol{vg} - \boldsymbol{g} - \boldsymbol{a} - \boldsymbol{b} - \boldsymbol{vb}$. The interpretation of the evaluation average was almost bad; if we want most answers to be very good and good, we admit not many to be average.

Therefore, in this case $\boldsymbol{J} = \boldsymbol{G} \cup \boldsymbol{B}$, $\boldsymbol{G} = \{\boldsymbol{vg}, \boldsymbol{g}\}$, $\boldsymbol{B} = \{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{vb}\}$; the statistical distributions (percentage histograms) of answers were interpreted in the sense of multiple criteria optimization, with $j \in \boldsymbol{G} = \{\boldsymbol{vg}, \boldsymbol{g}\}$ as quality indicators that should be maximized, and $j \in \boldsymbol{B} = \{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{vb}\}$ as quality indicators to be minimized.

A reference point approach was proposed for this particular case of ranking probability distributions; other approaches are usually more complicated (see, e.g., [17]). However, when the dean of the School of Knowledge Science in JAIST (Yoshiteru Nakamori) was asked to define his preferences or preferred aspiration levels, the reality of the managerial situation overcome his theoretical background: he responded "*in this case, I want the ranking to be as objective as possible – I must discuss the results with the deans of other schools and with all professors*". This was the origin of reflection on objective versus subjective rational ranking.

Thus, a statistical average of the percentages of answers in the entire data set was taken as the reference distribution or profile. Since such a reference profile might result in good but standard answers, some artificial reference distributions were also constructed as more demanding than the average one; averages over Pareto optimal options were not computed because of the complexity of the data set.

The detailed results of the survey were also very useful for university management (see [23]). It was found that seven questions of the first (assessment) type ranked as worst practically did not depend on the variants of ranking; thus, the objective ranking gave robust results as to the problems that required most urgent intervention by the university management. The best ranked questions of the second (importance) type were more changeable, only three of them consistently were ranked among the best ones in diverse ranking profiles. Moreover, a rank reversal phenomenon was observed: if the average reference distribution was used, best ranked were questions of rather obvious type, more interesting results were obtained when using more demanding reference profile.

Another possible application of the concept of objective ranking is the issue of detecting a significant event in a network (say, a failure of a link in a computer network). We can observe certain characteristic variables in the network and their histograms – empirical probability distributions. In the case of failure, these probability distributions will change as compared to the case of normal network functions; the issue is to use such change to identify the type of the event. Thus, the decision options $k \in \boldsymbol{K}$ in this problem are possible types of events; for each type of event, we might have a reference probability distribution, obtained, e.g., via network simulation. In such a case, the detection of the type of event is equivalent to checking which reference probability distribution is the closest to the actual empirical distribution; this can be done using also reference profile approach with stabilized criteria. However, another approach is to try to define a partial order in the space of histograms that would represent the given problem of event detection, and use an objective ranking approach to produce a ranking list of types of events, given an empirical histogram. This will be the subject of further studies (see [6]).

5. Conclusions

We discussed in this paper some aspects of the general issue of objectivity versus subjectivity, with the conclusion that objectivity is a higher value, similar to justice: it might be not fully attainable, but it is worth striving for. We have also shown that the reduction of objectivity to power and money, suggested by postmodern sociology of science, is not only based on superficial reductionism, but also contains logical errors.

We presented in this paper the issue of objective ranking defined as dependent only on a given set of data, relevant for the decision situation, and independent of any more detailed specification of personal preferences than that given by defining criteria and the partial order in criterion space. Rational objective ranking can be based on reference point approach, because reference levels needed in this approach can be established objectively statistically from the given data set.

Examples show that such objective ranking can be very useful in many management situations.

There are several possible topics for further study, such as the relation of objective ranking obtained by reference point approaches and objective ranking obtained by rough set approaches, since the latter also can be seen as dependent only on a given set of data, on an informational system in the sense of Zdzisław Pawlak (see [18] and [7]), or the issue of using multiobjective comparison of empirical statistical profiles for event detection in telecommunication networks [6].

References

[1] R. I. Ackoff, "Towards a behavioural theory of communication", in *Modern Systems Research for the Behavioural Scientist*, W. Buckey, Ed. Chicago: Aldine, 1957.

[2] L. Bertallanfy, "General systems theory", *Gener. Syst.*, vol. 1, pp. 1–10, 1956.

[3] M. Ehrgott and D. Tenfelde-Podehl, *Nadir Values: Computation and Use in Compromise Programming.* Universität Kaiserslautern Fachbereich Mathematik, 2000.

[4] M. Foucault, *The Order of Things: An Archeology of Human Sciences*. New York: Routledge, 1972.

[5] J. Granat, M. Makowski, and A. P. Wierzbicki, "Hierarchical reference approach to multicriteria analysis of discrete alternatives", in *CSM'06: 20th Worksh. Method. Tools Compl. Syst. Model. Integr. Policy Asses.*, Laxenburg, Austria, 2006.

[6] J. Granat and A. P. Wierzbicki, "Objective classification of empirical probability distributions and the issue of event detection" , in *23rd IFIP TC 7 Conf. Syst. Model. Optim.*, Cracow, Poland, 2007.

[7] S. Greco, B. Matarazzo, and R. Slowinski, "Rough sets theory for multicriteria decision analysis", *Eur. J. Oper. Res.*, vol. 129, no. 1, pp. 1–47, 2001.

[8] W. Heisenberg, "Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik", *Zeitschrift für Physik*, vol. 43, pp. 172–198, 1927 (in Germany).

[9] H. S. Jensen, L. M. Richter, and M. T. Vendelø, *The Evolution of Scientific Knowledge*. Cheltenham: Edward Elgar, 2003.

[10] R. Keeney, *Value Focused Thinking, a Path to Creative Decision Making*. Cambridge: Harvard University Press, 1992.

[11] R. Keeney and H. Raiffa, *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. New York: Wiley, 1976.

[12] M. M. Kostreva, W. Ogryczak, and A. Wierzbicki, "Equitable aggregations and multiple criteria analysis", *Eur. J. Oper. Res.*, vol. 158, pp. 362–377, 2004.

[13] B. Latour, *Science in Action*. Milton Keynes: Open University Press, 1987.

[14] L. Lessig, *Free Culture: The Nature and Future of Creativity*. London: Penguin Books, 2004.

[15] K. Miettinen, *Nonlinear Multiobjective Optimization*. Boston: Kluwer, 1999.

[16] W. Ogryczak, "On multicriteria optimization with fair aggregation of individual achievements", in *CSM'06: 20th Worksh. Method. Tools Compl. Syst. Model. Integr. Policy Asses.*, Laxemburg, Austria, 2006.

[17] W. Ogryczak and A. Ruszczyński, "On consistence of stochastic dominance and mean-semideviation models", *Math. Program.*, vol. 89, pp. 217–232, 2001.

[18] Z. Pawlak, *Rough Sets – Theoretical Aspects of Reasoning about Data*. Dordrecht: Kluwer, 1991.

[19] K. R. Popper, *Objective Knowledge*. Oxford: Oxford University Press, 1972.

[20] W. V. Quine, "Two dogmas of empiricism", in *Philosophy of Mathematics*, P. Benacerraf and H. Putnam. Englewood Cliffs: Prentice-Hall, 1953.

[21] J. Rawls, *A Theory of Justice*. Cambridge: Belknap Press, 1971.

[22] T. Saaty, *Decision Making for Leaders: The Analytical Hierarchy Process for Decisions in a Complex World*. Belmont: Lifetime Learning Publ., 1982.

[23] J. Tian, A. P. Wierzbicki, H. Ren, and Y. Nakamori, "A study on knowledge creation support in a Japanese Research Institute", in *Proceeding of First International Conference on Knowledge Science, Engineering and Management KSEM'06*. Berlin-Heidelberg: Springer Verlag, 2006, pp. 405–417.

[24] A. P. Wierzbicki, "On the completeness and constructiveness of parametric characterizations to vector optimization problems", *OR Spektrum*, vol. 8, pp. 73–87, 1986.

[25] A. P. Wierzbicki, "Technology and change: the role of technology in knowledge civilization", in *1th World Congr. IFSR*, Kobe, Japan, 2005.

[26] A. P. Wierzbicki, M. Makowski, and J. Wessels, *Model-Based Decision Support Methodology with Environmental Applications*. Dordrecht: Kluwer, 2000.

[27] A. P. Wierzbicki and Y. Nakamori, *Creative Space: Models of Creative Processes for the Knowledge Civilization Age*. Berlin-Heidelberg: Springer Verlag, 2006.

[28] A. P. Wierzbicki and Y. Nakamori, Eds., *Creative Environments: Creativity Support for the Knowledge Civilization Age*. Berlin-Heidelberg: Springer Verlag, 2007.

Andrzej Piotr Wierzbicki born June 29, 1937 in Warsaw. Graduated in 1960 as Master of Engineering at the Faculty of Telecommunications, Warsaw University of Technology (WUT). He took a Ph.D. degree at this University in 1964, for a thesis on nonlinear feedback systems; D.Sc. degree in 1968, for a thesis on optimization of dynamic systems. In 1971–75 Deputy Director of the Institute of Automatic Control, later Deputy Dean of the Faculty of Electronics, Warsaw University of Technology. In 1975–78 the Dean of the Faculty of Electronics, WUT. Since 1978 worked with the International Institute for Applied Systems Analysis in Laxenburg n. Vienna, Austria; 1979–84 as the chairman of the theoretical branch, Systems and Decision Sciences Program, of this Institute. From 1985 back in the Institute of Automatic Control, WUT, as a Professor of optimization and decision theory. In 1986–91 scientific secretary of the Committee of Future Studies "Poland 2000" (in 1990 renamed "Poland 2000+") of P.Ac.Sc.; currently deputy chairman. In 1991 elected a member of the State Committee for Scientific Research of Republic of Poland and the chairman of its Commission of Applied Research. Deputy chairman of the Council of Polish Foundation for Science (1991–94). Chairman of scientific councils of NASK (National Scientific and Academic Computer Network) in Poland (1994–2004) and PIAP (the Industrial Institute of Measurements and Control) (1992–2004). Editor in chief of the quarterly "Archives of Control Sciences" of P.Ac.Sc. (1986–96). Since 1997 deputy chairman of the Committee for Cooperation with IIASA of P.Ac.Sc. In 1996–2004 the Director of the National Institute of Telecommunications in Poland. In 2000–2003 a member of ISTAG (Information Society Technology Advisory Group) of European Commission. In 2001–2004 the chairman of the Advisory Committee on Scientific International Relations of the State Committee for Scientific Research of Poland. In 2004–2007 a Research Professor at National Institute of Telecommunications in Poland and at Japan Advanced Institute of Science and Technology, Nomi, Ishikawa, Japan. Beside teaching and lecturing for over 45 years and promoting over 80 master's theses and 18 doctoral dissertations at WUT, he also lectured at the Faculty of Mathematics of Warsaw University and at doctoral studies: at WUT; at the Academy of Mining and Metallurgy in Cracow; at the University of Minnesota; at the Technische Hochschule Ilmenau; at Fernuniversität Hagen; at Kyoto University. Author of over 180 publications, including 12 books (5 monographs, 7 – editorship of joint international publications), over 60 articles in scientific journals, over 90 papers at conferences; author of 3 patents granted and industrially applied. Current interests include parallelization of optimization algorithms using vector optimization and multiple criteria approaches, diverse aspects of negotiation and decision

support, including, e.g., applications of fuzzy set theory for describing uncertainty in decision support models, multimedia software in computer networks and in distance education, diverse issues of information society and civilization, rational theories of intuition and of knowledge creation. Languages: English, German, Russian – each fluent, beside native Polish. Member of IEEE, ISMCDM (International Society of Multiple Criteria Decision Making), SEP (Polish Society of Electrical Engineers), PTM (Polish Mathematical Society), PSKR (Polish Association for the Club of Rome). In 1992 received (as first European researcher) the Georg Cantor Award of the International Society of Multiple Criteria Decision Making for his contributions to the theory of vector optimization and multiple criteria decision support. In 2004 received Belgian Royal Distinction Le Merite de l'Invention, d'Officier Croix. In 2004 received Tomasz Hofmokl Prize of NASK of Poland for promoting the idea of informational society. In 2005 received Best Paper Award at the 38th Hawaii International Conference of System Science, for the paper "Knowledge creation and integration: creative space and creative environments" (with Y. Nakamori).
e-mail: A.Wierzbicki@itl.waw.pl
National Institute of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland

Paper

Objective classification of empirical probability distributions and the issue of event detection

Janusz Granat and Andrzej P. Wierzbicki

Abstract—The paper concentrates on the issue of classification of empirical probability distributions (histograms), which is useful both in management situations and in event detection or event mining. While existing approaches to event detection concentrate on the use of selected moments or other characteristics of empirical probability distributions, we postulate that full empirical distribution preserves more of needed information then selected moments of this distribution, thus multiple criteria classification of distributions can be most effective in event detection.

Keywords— objective classification, event detection, multicriteria analysis.

1. Introduction

While practically all multiple criteria approaches to decision analysis and support concentrate on rationally supporting *subjective* decisions, depending on some form of an elicitation of preferences of the decision maker, there are diverse decision situations where we should suggest decisions that are made *as objectively as possible*; the full objectivity is not attainable for many practical and philosophical reasons, but objectivity can be seen as an useful ideal or goal. Examples of such situations are, on the one hand, managerial decisions influencing many stakeholders, when an aggregation of preferences of stakeholders is impossible. On the other hand, such situations occur also in event detection, e.g., when automatically detecting a case of fire, we should not make decisions based on subjective, personal preferences.

We shall call the problem of supporting decisions in such a case the problem of *objective classification* (treating problem of *ranking* as a special case with singleton classes and the problem of *decision selection* and *detection* as special cases with classes *selected – not selected* or *detected – not detected*). We can define objective classification as dependent only on a given set of data, relevant for the decision situation, and independent from any more detailed specification of personal preferences than that given by defining criteria and the partial order in criterion space. Already in this definition, we see the limits to objectivity, because naturally the definition of criteria and their partial order, or of the relevant set of data, can be treated as subjective; however, they are often much more obvious and easy to agree upon than the detailed preferences defined, e.g., by a utility function or a set of weighting coefficients.

Most of classical approaches to multiple criteria decision analysis and support, e.g., based on weighted sum aggregation, are not easily adaptable to the case of objective classification. From known approaches, either the goal programming or the reference point approaches are easily adaptable, because goals or reference points can be defined reasonably objectively from statistics in a given set of data. We concentrate here on reference point approaches, because they have the property of producing always Pareto optimal options (which is not the case in goal programming).

In this paper we are focusing on the issue of classification of empirical probability distributions (histograms), which is useful both in management situations and in event detection or event mining. While existing approaches to event detection concentrate on the use of selected moments or other characteristics of empirical probability distributions, we postulate that full empirical distribution preserves more of needed information then selected moments of this distribution, thus multiple criteria classification of distributions can be most effective in event detection. One of advantages of reference point approaches is that they easily deal with so-called *multiobjective trajectory* analysis and optimization; this can be applied to issues of stochastic dominance and their generalizations needed for multiple criteria event detection based on classification of empirical probability distributions.

2. An algorithm for event detection

An outline of an algorithm for event detection based on histograms is as follows:

1. Calculate a typical histogram (based, e.g., on historical data):

   H_{ok} – the value of a typical histogram for an interval k, $k = 1, \ldots, N$.

2. Identify a priori the set of events E and(or) corresponding anomalies A:

   e_i – an event i, $i = 1, \ldots, I$,

   A_i – an anomaly corresponding to event i, $i = 1, \ldots, I$.

3. Identify the set of characteristic histograms for the set of events (using, if necessary, simulation of events and anomalies):

   H_{ik} – a histogram characteristic for event or anomaly i,

   while $i = 0$ corresponds to lack of anomaly (the typical histogram).

4. Define partial orders corresponding to each event (whether an event typically results in a decrease or increase or lack of change of typical histogram values, see below):

   PO_i – a partial order of histograms.

5. Calculate the histogram of the observed current measurements from the real system:

   $H_k, \quad k = 1, \ldots, N$ – the histogram of observed measurements.

6. Calculate achievement values comparing the histogram of observed measurements with histograms of the typical character and of character characteristic for specific events, treated as $I+1$ $(i = 0, \ldots, I)$ different reference trajectories.

7. Detect an event based on achievement values (by selecting the event with corresponding highest achievement value).

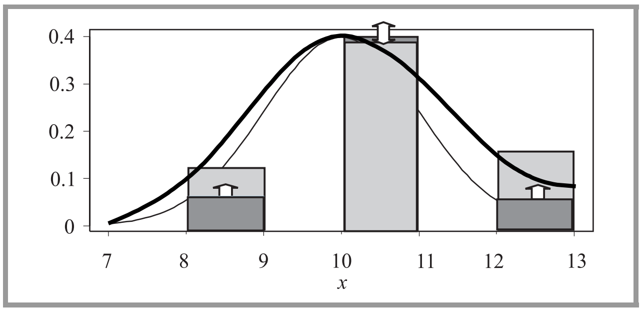


Fig. 1. Partial order of histograms.

A partial order of histograms (Fig. 1) is defined as follows. Two histograms H_α, H_β (with values $H_{\alpha,k}$, $H_{\beta,k}$ satisfy a given partial order $PO_i(H_\alpha, H_\beta)$, if:

$$PO_i(H_\alpha, H_\beta)$$

$$H_{\alpha,l} \leq H_{\beta,l}, \qquad l = 1, \ldots, L$$

$$H_{\alpha,k} \geq H_{\beta,k}, \qquad k = 1, \ldots, K$$

$$H_{\alpha,m} \approx H_{\beta,m}, \qquad m = 1, \ldots, M$$

$$K + L + M = N,$$

where:

N – the number of intervals (common for both histograms).

Achievement values can be defined as follows (for the case $K = N$, $L = M = 0$; for other cases see, e.g., [8]):

$$\delta_i(H_k, H_{ik}) \,|\, PO_i = \min_k (H_k - H_{ik}) + \varepsilon \sum_k (H_k - H_{ik}),$$

$$i = 0, \ldots, I$$

Let us consider a very simple example (see Fig. 2). In the top left side of Fig. 2 we have two histograms H_{O1}, H_{O2} that represent histograms for normal operation of the system measured at two different outputs. However, in the bottom left side we have actually observed histograms H_1, H_2 that differ substantially from normal operation. In the right hand side of the picture we have defined histograms for two events e_1, e_2, with arrows indicating partial orders. Do the actually observed histograms correspond to normal operation, event e_1 or e_2?

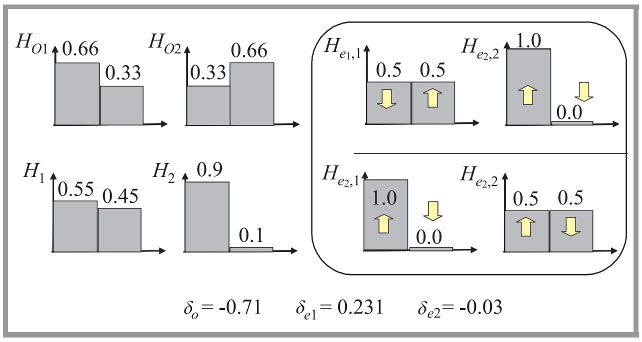


Fig. 2. An event detection.

If we calculate values of the achievement functions for the given histograms H_1, H_2 that result from the actual measurements, using normal operation histogram and the histograms characteristic for the two events as three different reference trajectories, we obtain:

$$\delta_o = -0.71, \quad \delta_{e1} = 0.231, \quad \delta_{e2} = -0.03\,.$$

Therefore, we can identify event 1, having the maximal achievement value, as the one best representing the measurements and the assumed partial order.

3. Fault detection

In this section we will discuss how the presented approach can be applied to detection of faults in a computer network.

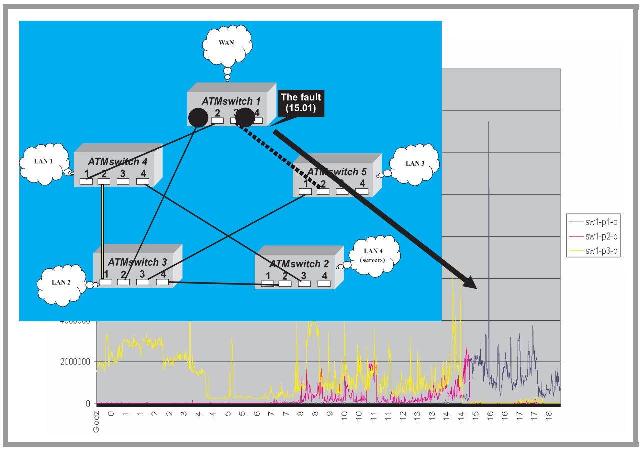


Fig. 3. A fault of the computer network.

Figure 3 shows a network with five switches. On 16.01 we can observe a fault of the connection between switch 1 and switch 5. The observations of the network throughput are presented on Figs. 4 and 5.

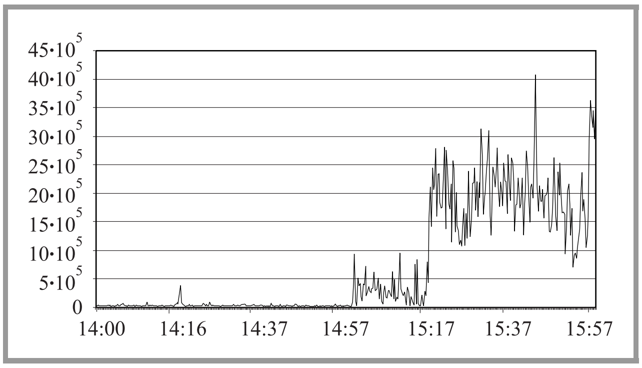


Fig. 4. Observation of the network throughput – switch 1, port 1.

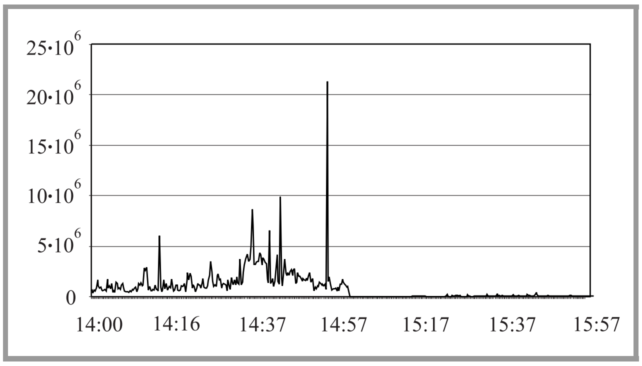


Fig. 5. Observation of the network throughput – switch 1, port 3.

We can transform this observations to the histograms shown of Figs. 6 and 7.

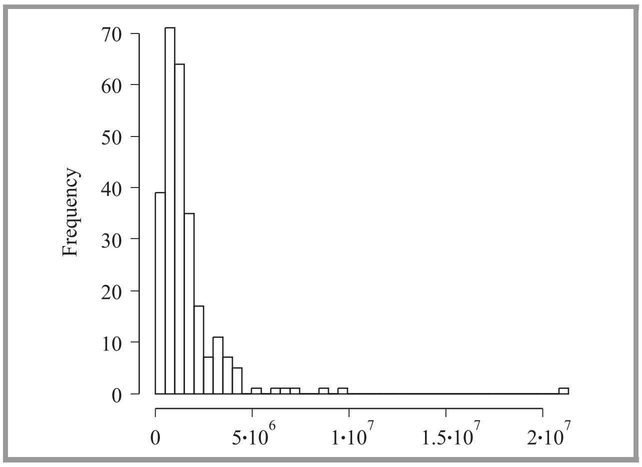


Fig. 6. A histogram of the observation on switch 1, output 3 – normal state of operation.

We see on Figs. 6 and 7 that the number of intervals that are significant from the point of event detection is small. Only these few intervals can be considered in the process of event detection.

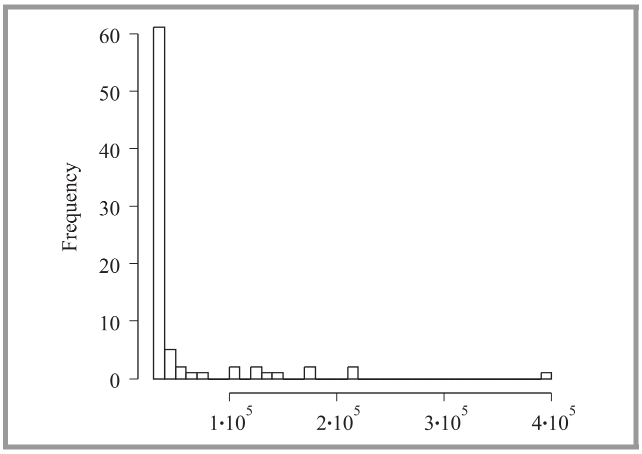


Fig. 7. A histogram of the observation on switch 1, output 3 – after fault of the network.

4. Conclusions

The paper presents a new concept of event detection that is based on histograms and multicriteria approach. This approach allows to consider only selected intervals of the histograms what is especially important in the problem of event detection where the significant changes can be detected only in selected intervals. It is also possible to consider histograms corresponding to observations of various points of the system. In Section 2 we have presented one of possible algorithms for event detections; other variants of the presented algorithm are also possible. The disadvantage of this approach is that we have to specify a priori the given sets of events and the histograms characteristic for given events. The advantage is that we use in a possibly most complete way the statistical information contained in the measurements.

References

[1] J. Granat, "Events driven approach for supporting real-time management of complex systems", in *Proc. First Word Congr. Int. Feder. Syst. Res. IFSR*, Kobe, Japan, 2005.

[2] J. Granat, "Event mining based on observations of the system", *J. Telecommun. Inform. Technol.*, no. 3, pp. 87–90, 2005.

[3] J. Granat and A. P. Wierzbicki, "Multicriteria analysis in telecommunications", in *Proc. 37th Ann. Hawaii Int. Conf. Syst. Sci. HICSS'04*, Big Island, Hawaii, USA, 2004.

[4] J. Granat, "Multicriteria analysis for behavioral segmentation", *J. Telecommun. Inform. Technol.*, no. 3, pp. 39–43, 2004.

[5] D. Luckham, *The Power of Events: An Introduction to Complex Event Processing in Distributed Enterprise Systems*. Boston: Addison-Wesley, 2002.

[6] L. Perrochon, W. Mann, S. Kasriel, and D. C. Luckham, "Event mining with event processing networks", in *Pacific-Asia Conference on Knowledge Discovery and Data Mining, LNCS*. London: Springer, 1999, vol. 1574, pp. 474–478.

[7] A. P. Wierzbicki, "On the completeness and constructiveness of parametric characterizations to vector optimization problems", *OR Spektrum*, vol. 8, pp. 73–87, 1986.

[8] A. P. Wierzbicki, M. Makowski, and J. Wessels, *Model-Based Decision Support Methodology with Environmental Applications*. Dordrecht: Kluwer, 2000.

Janusz Granat received his M.Sc. in control engineering (1996) and his Ph.D. (1997) in computer science from the Warsaw University of Technology, Poland. He holds a position as an Assistant Professor at the Warsaw University of Technology, and is the leader of a research group on applications of decision support systems at the National Institute of Telecommunications in Warsaw. He lectured decision support systems and various subjects in computer science. His scientific interests include data mining, modeling and decision support systems, information systems for IT management. Since 1988 he has been cooperating with IIASA. He contributed to the development of decision support systems of DIDAS family and the ISAAP module for specifying user preferences. He has been involved in various projects related to data warehousing and data mining for telecommunication operators. He was also involved in EU MiningMart project.

e-mail: J.Granat@itl.waw.pl
National Institute of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland

Institute of Control and Computation Engineering
Warsaw University Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland

Andrzej P. Wierzbicki – for biography, see this issue, p. 22.

Paper

Supporting telecommunication product sales by conjoint analysis

Piotr Rzepakowski

Abstract—Conjoint analysis is widely used as a marketing research technique to study consumers' product preferences and simulate customer choices. It is used in designing new products, changing or repositioning existing products, evaluating the effect of price on purchase intent, and simulating market share. In this work the possibility of conjoint analysis usage in telecommunication filed is analyzed. It is used to find optimal products which could be recommended to telecommunication customers. First, a decision problem is defined. Next, the conjoint analysis method and its connections with ANOVA as well as regression techniques are presented. After that, different utility functions that represent preferences for voice, SMS, MMS and other net services usage are formulated and compared. Parameters of the proposed conjoint measures are determined by regression methods running on behavioral data, represented by artificially generated call data records. Finally, users are split in homogenous groups by segmentation techniques applied to net service utilities derived from conjoint analysis. Within those groups statistical analyses are performed to create product recommendations. The results have shown that conjoint analysis can be successfully applied by telecommunication operators in the customer preference identification process. However, further analysis should be done on real data, other data sources for customer preference identification should be explored as well.

Keywords— decision analysis, multiple criteria analysis, utility theory, preference measurement, conjoint analysis, consumer behavior, purchase intent, marketing, marketing tools.

1. Introduction

Selling is a practical implementation of strategies derived from marketing. One of them is loyalty management approach that is commonly used by telecommunication companies. Usually, loyalty programs are organized for but to remain competitive on deregulated market, other tasks like product recommendation should be done to maximize customer satisfaction. People who are satisfied with product usage are also loyal, since they do not need to change product supplier.

There is a permanent price reduction of telecommunication services and new products are launched so often that customers are not able to analyze all possibilities regularly and find the best products for themselves. Therefore, methods for preference identification should be developed to support telecommunication operator consultants with tools for products and services recommendation. In this work we have used conjoint analysis (CA) method to identify preferences of telecommunication customers. Contrary to the original method, instead of making a questionnaire, behavioral data were used to find real preferences not declared ones.

The paper is organized as follows. In Section 2, the basic requirements and customer preferences are reviewed. In Section 3, the optimization problem is explained and methods for problem solving are introduced. Also some assumptions are made to decrease the complexity of the problem. In Section 4, conjoint analysis process is described: the preference function is proposed and the statistical model for preference identification is created. Results are shown in Section 5 and conclusions are made in Section 6.

2. Preliminaries

2.1. Loyalty

Deregulation brought new competition that forces telecommunication companies as well as other retailers to implement new sale strategies. Boston Consulting Group indicates ten quality drivers [2] that should be addressed to remain competitive:

- call center,
- complaint management,
- customer communication,
- offer development,
- branding,
- sales channels,
- customer understanding,
- loyalty,
- e-utility.

Good practices where divided into three stages: "mastering the basics", "rising the bar" and "changing the game". At the beginning "mastering the basics stage" the first four dimensions are most important. At the second stage offer development, branding and sales channels are essential. However, in deregulated, full competitive markets deep customer understanding, loyalty and e-utility must be addressed to have a real chance in the competition.

The role of loyalty is increasing owing to wide range of benefits. Loyal customer:

- provides positive advertising through his recommendations to family and friends,
- is more receptive to cross-selling,
- provides company with feedback,
- tends to be more profitable.

2.2. Customer preferences

Because customer's loyalty depends on the satisfaction he gets from product and service usage, delivered goods should not only be of good quality but also should be well suited to user requirements. That leads marketing departments to research activity for identification and anticipation of clients' needs. Research results are used to design new products and deliver attractive goods to consumers. Attractive products are those which have sufficient functionality and acceptable price. Therefore, products of lower price and suited to user needs should be recommended. Paradoxically, telecommunication companies should take care about customers recommending cheapest products of that which are functionally acceptable. Customers should be sure that they do not pay extra money for not used additional features.
Two customer preference groups can be distinguished: real preferences and declared ones. Real preferences can be derived directly from information about bought products, services and product usage. However, that information tells us only about past preferences and is limited to existing products functionality. Declared wishes gathered from questionnaires, contrary to real preferences, give additional information about future needs and are not restricted to existing product functionality but do not have to correspond to real ones. Differences in declared and real requirements are caused by uncertainty of results obtained from questionnaires but also by limited information represented in behavioral data. To analyze preferences entirely both real and declared preferences should be considered. Additionally, preferences can also be derived indirectly from demographic, geographic and socioeconomic data connected with user behavior or declared needs. Nevertheless, in this article we restrict analysis to real preferences obtained from behavioral data.

2.3. Telecommunication products

Telecommunication customers pay for net services (voice, short message service (SMS), multimedia messaging service (MMS) and general packet radio service (GPRS)) usage. However, price for services is dissimilar for them. Cheaper services are for users who declare to use services in fixed period or in minimum amount. For example, people who signed contracts have cheaper calls then the others. Cost can be also reduced by additional packages valid for a short period or other products that can be used only in specific time. For example, there are usually accessible packages that reduce call price after working hours or in the weekends. Those additional packages will be called further telecommunication products.
Products are provisioned at the end user level or at the account level. End user is associated with the account and is rated for service usage in the way defined by tariff plan he has. Some products are allowed at discount prices if there were bought other services or products. Moreover, more then one user can be associated with the account. Therefore, if products are installed at the account level they can be used by many users. For all users pays owner of the account who is called customer. Furthermore, there are additional businesses constraints that make some products unavailable at various tariff plans and some products are switched off which means that new installations cannot be made any more.
That is a big challenge for customers to be on time with all promotions and to find the most fitted products for all users on the account. That task requires identification of users' needs and solving a complex optimization problem.

2.4. Conjoint analysis

In this article, usage of conjoint analysis technique is proposed for customers' preference identification. CA is well known in marketing research field and is commonly used to identify consumer preferences from a questionnaire data. It provides preferences in compact form as parameters of the utility function a priori defined by an analyst. CA allows determining relative preference structure that can be easily used to compare clients, make segmentation and profiling. When all of the attributes are nominal, the metric conjoint analysis is a simple main-effect analysis of variance (ANOVA) with some specialized output. The ANOVA problem can be solved using regression techniques what is shown in Section 4.

3. Service costs optimization problem

The task is to find optimal set of products individually for each user. Usually a telecommunication operator has dozen million users and more then one hundred products in an offer. Because business constrains complicate the problem some assumptions are made to simplify it.

3.1. Business constraints

Three main groups of business constraints can be distinguished:

1. Tariff plan constraints:
   - user can change the tariff plan he has to a higher one then the one he signed in the contract;
   - old tariffs cannot be used any more.
2. Product constraints:
   - some old products cannot be sold any more;
   - only a few additional products are allowed within particular tariff plan;
   - usage of some products excludes usage of others;

- in some tariff plans users have to choose the demanded quantity of products;
- value of some packages can be defined individually;
- products can be defined on end user level as well as at the account level.

3. Product usage constraints:
   - data are accessible monthly;
   - some users do not make enough connections monthly to analyze the data;
   - users are charged in billing cycles that start on different days of month;
   - to have all information about connections for new users there is a need to analyze at least two months of data;
   - some products can be activated with a delay, for example from the customer billing cycle date.

3.2. Assumptions

To simplify the problem optimal products for the end user in his actual tariff plan will be found. Instead of optimal set of products, ranking lists of them will be made using only two months history of outgoing calls.

1. There would only be analyzed services within current users' tariff plans. Changes of tariff plans are not under consideration in this work. Tariff plans can be also treated as other services but to do so additional business knowledge about configuration constraints is required.

2. There would be created recommendation lists of services at the end user level. Also services that cannot be sold any more would be recommended. If some of them cannot be sold or are not allowed in current user tariff plan they will be removed later after creation of the ranking. The removal of services depends only on business constraints and is not taken into consideration in this work.

3. Data from two months will be analyzed and users who make less then 50 calls will be removed, since there is no need to sell them additional products.

4. Only outgoing calls will be analyzed because products reduce only those costs.

3.3. Optimization problem

Indicies:

s – end user,

p – product,

a – attribute.

Parameters:

S – finite and nonempty set of end users,

P – finite and nonempty set of products,

P_s – finite set of products that are available for users,

D_s – finite and nonempty set of call data records (CDR) from one billing cycle of customers,

A – finite and nonempty set of CDR attributes,

$V = \cup_{a \in A} V_a$, V_a is a set of values of attribute a, called the domain of a,

F – finite and nonempty set of rating function definitions for each product p,

C – finite and nonempty set of products' orders.

Decision variables:

x_s – finite set of customer products.

Constraints:

$s \in S$,

$x_s \subseteq P_s \subseteq P$,

$a \in A$.

Functions:

ρ – rating function.

Objective value:

$$\min_{x_s} \rho(D_s, x_s, F, C) \quad \forall s \in S. \tag{1}$$

3.4. Optimization methods

Decision problem described in the previous section can be solved using:

- optimization techniques,
- simulation techniques or
- statistical analysis and data mining techniques.

Optimization and simulation methods give very good results but are very slow and resource consuming. Checking all combinations of products for dozen million of users would require as much resources of rating infrastructure as telecommunication operator possess multiplied by number of possible product combinations. Simulation is impossible in practice because of huge maintenance costs. On the contrary, optimization techniques are usually faster but also are time and resource consuming. Furthermore, both of those methods require precise knowledge about rating functions F defined for each product and cascade definitions C to apply functions correctly. Often knowledge about those functions is distributed between systems and functions are represented in different ways dedicated for tool and specially formatted data. Costs of data collection and algorithm standardization are very large and in consequence increase maintenance costs of optimization and simulation models to unacceptable level.

Statistical analysis and data mining techniques are less accurate then methods described earlier. However, the precise knowledge about rating functions is not needed and can be practically applicable for huge amount of data. Instead of rating knowledge they use statistical information about cost of service for different users who have installed different products. It is assumed that most of people buy products in order to reduce cost of telephone usage. Only some of them do not have time or they do not want make analysis to find the best products. Nevertheless, people who behave similarly should have analogous sets of products. Thus, the idea is to find similar users and recommend them products which are used most frequently in their group.
We decided to use clustering method and statistical analysis to solve the decision problem. Usually, in model creation process some transformations are performed on input data [8]. We add customer preference identification step to improve analysis. In that additional step conjoint analysis for preference identification is used.
Summing up, there are three main tasks to recommend products:

– user needs identification by conjoint analysis,
– user clustering,
– statistical analysis.

4. Conjoint analysis for customer preference identification

Conjoint analysis process consists of [16]:

– selection of utility factors,
– conjoint measure definition,
– conjoint model definition,
– questionnaire preparation,
– questionnaire data acquisition,
– statistical analysis,
– data interpretation.

For utility factors we get some attributes from behavioral data. The questionnaire preparation step is not required because historical data is analyzed. Hence, the questionnaire data acquisition step changes to the behavioral data preparation one.

4.1. Selection of utility factors

Attributes which differentiate the cost of services most were chosen as utility factors. Among them there are: service, location, network, and day types with categories presented in Table 1. Original call data records were transformed to determine chosen attributes. Next, data is aggregated and statistics of call frequencies for each aggregation were calculated.

Table 1
Utility factors

Attribute	Levels
Service	Voice
	SMS
	MMS
	GPRS
Location	Home
	Roaming
Net	To on-net
	To off-net (mobile operators)
	To fixed (fixed operators)
	To international (international operators)
Day type	Working days
	Weekend or holiday

4.2. Conjoint measure definition

The dependency between utility factors is defined by the conjoint measure. It consists of intercept coefficient μ and part-worth utilities associated with attributes A. If some attributes are correlated then the interaction between those attributes are added to the conjoint measure. Interactions between pairs are usually enough but sometimes interactions of higher types, for example between three variables are used. An example of conjoint measure defined for three attributes is presented in Eq. (2):

$$\begin{aligned} y &= \mu + \alpha_{A_1} + \alpha_{A_2} + \alpha_{A_3} \\ &+ \beta_{A_1A_2} + \beta_{A_1A_3} + \beta_{A_2A_3} \\ &+ \gamma_{A_1A_2A_3} + \varepsilon \,. \end{aligned} \tag{2}$$

In that example part worth utilities are presented by α vectors of utilities for attribute values, β vectors of utilities for all combinations of values associated with two attributes and γ vector of utilities for combination of values taken from attribute A_1, A_2 and A_3. If all values of part-worth utilities are known then utility value for each call can be counted.
To make the results unique the equation must also fulfill conditions:

$$\sum_{v_a \in V_a} \alpha_{a v_a} = 0, \quad \forall a \in A\,, \tag{3}$$

$$\sum_{v_a \in V_a} \beta_{a b v_a v_b} = 0, \quad \forall a \in A, \forall b \in A, a \neq b, \forall v_b \in V_b\,, \tag{4}$$

$$\sum_{v_b \in V_b} \beta_{a b v_a v_b} = 0, \quad \forall a \in A, \forall b \in A, a \neq b, \forall v_a \in V_a\,. \tag{5}$$

A similar condition for γ parameters has to be defined.
For presented telecommunication task we have compared two measures. One of them consisted of linear terms and

correlation between all pairs of attributes. Another one was extended by interactions between three attributes. After the analysis, the second measure with factors presented in Table 2 has been chosen.

Table 2
Conjoint measure factors

Attribute	Levels
Service	4
Location	2
Net	4
Day type	2
Service*location	8
Service*net	16
Service*day type	8
Location*net	8
Location*day type	4
Net*day type	8
Service*net*day type	32
Total	96

Finally, conjoint measure presents Eq. (6):

$$\begin{aligned} y \;=\; & \mu \\ & + \alpha_{service} + \alpha_{location} + \alpha_{net} + \alpha_{day\,type} \\ & + \beta_{service*location} + \beta_{service*net} \\ & + \beta_{service*day\,type} + \beta_{location*net} \\ & + \beta_{location*day\,type} + \beta_{net*day\,type} \\ & + \gamma_{service*net*day\,type} \\ & + \varepsilon\,. \end{aligned} \quad (6)$$

4.3. Conjoint model definition

Conjoint model is a statistical model that represents dependencies between utility of a profile and its attributes and is defined by Eq. (7):

$$y = \alpha^T x + \varepsilon\,. \quad (7)$$

Now α coefficient represent utilities associated with all conjoint factors α, β and γ defined earlier. Because all of attributes of conjoint measure are categorical, dummy variables x created to represent no metric information. One attribute with k levels was replaced by $k-1$ binary attributes.

After adding dummy variables regression techniques can be used for part worth utilities identification. Dependant variable y in the regression model represents utility of a profile. In analyzed problem it was calculated as probability of making a call which means that it has binomial distribution. That problem cannot be solved simply by linear regression as regression techniques required normal distribution of dependant variable. However, binomial distribution can be simply transformed to the normal one by logit function. In consequence, general linear model (GLM) was defined as

$$\ln\left(\frac{y}{1-y}\right) = \alpha^T x + \varepsilon\,, \quad (8)$$

$$y = \frac{e^{\alpha^T x + \varepsilon}}{1 - e^{\alpha^T x + \varepsilon}}\,. \quad (9)$$

5. Analytical results

5.1. Conjoint analysis

To make analysis we generated artificial CDR for 1000 users. The data were transformed in statistical analysis software (SAS) to prepare full profile ranking lists. Using SAS procedure TRANSREG [12] conjoint model was fitted individually for each user. The attributes were automatically coded to binary variables by that procedure. As a result we get relative importance of the attributes for each user and the part worth utilities connected with attribute values. The relative importance of each attribute was calculated from the utilities of attributes as [16]

$$I_a = \frac{\max_{v_a}\{U_{av_a}\} - \min_{v_a}\{U_{av_a}\}}{\sum\limits_{a\in A}(\max_{v_a}\{U_{av_a}\} - \min_{v_a}\{U_{av_a}\})}\,, \quad (10)$$

where:

U_{av_a} – part worth utility associated with v-value of a-attribute,

v_a – value of attribute a.

Analytical results are presented for two models:

- logit II: GLM model with logit transformation on dependant variable, linear term and interactions between all attribute pairs;
- logit III: logit II + the interaction of three variables: service, net and day type.

Comparison of average relative importance of conjoint model attributes and standard deviation statistics for two models are presented in Table 3. The service*net*day type attribute is quite significant in the model and statistical tests confirm that all coefficients are significantly greater then zero. However, standard deviations have similar values to averages what means that user groups are not homogenous. People in population behave differently: use different services, prefer different nets and make calls in different days.

Statistics presented in Table 4 shows that both logit II and logit III models are well filled to data. Average value of R^2 is 99% and standard deviation is very low. The worst

Table 3
Relative importance statistics in population

Attribute	Logit II		Logit III	
	avg	std	avg	std
Service	21.0	15.2	20.1	14.9
Location	1.0	5.0	1.0	4.9
Net	22.8	14.4	22.1	14.2
Day type	13.9	10.5	13.0	10.5
Service*location	0.9	4.8	0.9	4.6
Service*net	18.8	15.9	17.4	15.5
Service*day type	8.5	8.2	5.2	7.8
Location*net	0.5	3.2	0.5	3.1
Location*day type	0.4	2.9	0.5	2.8
Net*day type	12.3	10.2	10.3	10.5
Service*net*day type	.	.	9.2	10.4

Table 4
ANOVA table

Model	Logit II	Logit III
min R^2	0.47	0.64
max R^2	1.00	1.00
avg R^2	0.99	0.99
std R^2	0.02	0.01
avg $adj-R^2$	0.89	0.81
std $adj-R^2$	0.21	0.30
avg p-value	0.17	0.22
std p-value	0.15	0.18

logit II model explains 47% of dependency in data and the worst logit III explains 67% of dependency in data. For further analysis logit III model has been chosen.

5.2. Customer clustering and product recommendations

Analytical results show that all users do not create homogenous group and recommendations of products cannot be made, yet. To find users with similar preference structure we have used results of conjoint analysis. Preference structure is defined by part worth utilities which have been calculated for each user individually using conjoint analysis methods. Now those coefficients can be used to make users clustering.

There are two types of clustering: hierarchical clustering and partition clustering. Hierarchical clustering proceeds successively by either merging smaller clusters into larger ones, or by splitting larger clusters. Partition clustering, on the other hand, attempts to directly decompose the data set into a set of disjoint clusters. For huge amount of data hierarchical clustering is not practically applicable, thus we used partition clustering implemented in SAS as a FASTCLUS procedure. In partition clustering number of clusters has to be given as an input to the procedure. There are different strategies to choose value which gives homogenous groups. As clustering methods are not under consideration of this work, 5 clusters were chosen to show the methodology.

Table 5
Average relative importance of attributes in segments [%] (logit III)

Attribute/segment	1	2	3	4	5
Service	19.4	13.9	30.4	26.4	15.8
Location	0.5	1.6	0.6	10.1	0.7
Net	20.9	13.1	23.7	18.1	21.8
Day type	9.7	3.1	10.1	5.6	14.6
Service*location	0.4	0.8	0.8	1.0	0.9
Service*net	32.1	8.5	14.2	5.8	18.9
Service*day type	3.7	6.1	5.6	5.0	5.2
Location*net	0.3	1.1	0.3	1.4	0.5
Location*day type	0.2	1.2	0.3	1.5	0.5
Net*day type	7.0	18.6	7.0	17.1	11.3
Service*net*day type	6.2	32.0	7.1	8.1	10.0

The results on average importance are presented in Table 5 and standard deviation statistics are illustrated in Table 6. The results show that users from those 5 segments behave differently. In the first segment service*net factor is mostly important (32%) while in the second segment service and net are correlated with day type and that coefficient is the most significant (32%). In other groups correlations of service and day type are much lower.

Table 6
Standard deviation of relative importance of attributes in segments [%] (logit III)

Attribute/segment	1	2	3	4	5
Service	11.7	8.4	11.7	13.1	14.1
Location	3.2	5.1	3.3	14.1	4.1
Net	9.4	10.0	10.8	13.3	15.4
Day type	5.3	5.5	6.1	6.9	11.7
Service*location	2.9	3.6	4.2	4.4	4.8
Service*net	10.8	9.9	11.2	9.6	16.6
Service*day type	5.6	9.6	6.8	7.8	8.2
Location*net	2.2	4.3	2.3	5.2	3.3
Location*day type	1.7	3.9	2.0	5.1	3.0
Net*day type	6.7	10.5	6.8	11.0	11.3
Service*net*day type	6.0	12.6	7.4	10.3	11.2

Standard deviations of relative importance are lower than in the whole population but are still comparable with average values of importance and further clustering should be done to divide presented groups into subgroups. The process should be repeated iteratively while users within groups have different preference structures. After getting homogenous groups, information about products can be added to each user and statistics can be made in those groups to find most frequently used services. Those services should be recommended to outliers who had bought different services then those which are most frequently used.

6. Conclusions and future research

Motivation and the use of conjoint analysis in telecommunication field were presented in this paper. The decision problem of finding optimal set of products for customers was defined and possible attitudes to solving the problem were compared. Conjoint analysis methodology and connections with ANOVA as well as regression techniques were presented. At the end, an example of preference identification process was introduced. Although, results from the example have shown that defined model explains dependency in data and in consequence customers' preference structures are accurate, further experiments on real data should be made. Also, additional information about users should be added including information about their declared preferences. Declared preferences might be quite interesting as with comparison to real ones they can indicate optimal actions which would allow increasing customers' satisfaction [3] and their loyalty at the same time.

References

[1] P. Aggarwal and R. Vaidyanathan, „Eliciting online customers' preferences: conjoint vs. self-explicated attribute-level measurements", *J. Market. Manage.*, vol. 19, pp. 157–177, 2003.

[2] J. Bissett, F. Falschini, Y. Jansen, R. Monti, and V. Massow, „Sparking connections: best practice in customer relationships and retail marketing", 2000, http://www.bcg.com

[3] P. J. Danaher, „Using conjoint analysis to determine the relative importance of service attributes measured in customer satisfaction surveys", *J. Retail.*, vol. 73, no. 2, pp. 235–260, 1997.

[4] W. S. DeSarbo *et al.*, „Representing heterogeneity in consumer response models", *Market. Lett.*, vol. 8, no. 3, pp. 335–348, 1997.

[5] C. Douglas and P. E. Green, „Psychometric methods in marketing research: part I, conjoint analysis", *J. Market. Res.*, vol. 32, no. 4, pp. 385–391, 1995.

[6] C. Douglas and P. E. Green, „Psychometric methods in marketing research: part II, multidimensional scaling", *J. Market. Res.*, vol. 34, no. 2, pp. 193–204, 1997.

[7] T. Evgeniou, C. Boussios, and G. Zacharia, „Generalized robust conjoint estimation", *Market. Sci.*, vol. 24, pp. 415–429, 2005.

[8] J. Granat, „Data mining and complex telecommunications problems modeling", *J. Telecommun. Inform. Technol.*, vol. 3, pp. 115–120, 2003.

[9] P. E. Green, S. M. Goldberg, and M. Montemayor, „A hybrid utility estimation model for conjoint analysis", *J. Market.*, vol. 45, no. 1, pp. 33–41, 1981.

[10] P. E. Green and A. M. Krieger, „Individualized hybrid models for conjoint analysis", *Manage. Sci.*, vol. 42, no. 6, pp. 850–868, 1996.

[11] P. E. Green, A. M. Krieger, and Y. Wind, „Thirty years of conjoint analysis: reflections and prospects", *Interfaces*, vol. 31, no. 3, part. 2, pp. 56–73, 2001.

[12] W. F. Kuhfeld, „Conjoint analysis", 2007, http://support.sas.com/techsup/technote/ts722h.pdf

[13] R. Johnson, http://www.sawtoothsoftware.com/education.shtml

[14] V. Srinivasan, „A conjunctive-compensatory approach to the self explication of multiattributed preferences", *Decis. Sci.*, vol. 19, no. 2, pp. 295–305, 1988.

[15] O. Toubia, J. R. Hauser, and D. I. Simester, „Polyhedral methods for adaptive choice-based conjoint analysis", *J. Market. Res.*, vol. 16, pp. 116–131, 2004.

[16] M. Walesiak and A. Bak, *Conjoint analysis w badaniach marketingowych*. Wrocław: WAE, 2000 (in Polish.

Piotr Rzepakowski received the M.Sc. degree in computer science from the Warsaw University of Technology, Poland, in 2003. Currently he is a Ph.D. candidate in computer science at the Warsaw University of Technology (Institute of Control and Computation Engineering). He is employed by National Institute of Telecommunications in Warsaw. Has taken part in projects related to data warehousing and analysis for a telecommunication operator. His research focuses on modeling, decision support, data mining, and customer preferences identification issues.

e-mail: Piotr.Rzepakowski@elka.pw.edu.pl
Institute of Control and Computation Engineering
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland

e-mail: P.Rzepakowski@itl.waw.pl
National Institute of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland

Paper

From transocean routes to global networks: a framework for liner companies to build service networks

Xiaoning Shi and Stefan Voß

Abstract—During the previous two decades liner carriers within maritime shipping have broken the barriers from being pure transportation providers towards being logistics service providers. Most of the top 20 liner carriers worldwide have set up spin-offs providing services from local booking up to 3rd party logistics services, combining the business advantages of tight linkages with liners together with the negotiation freedom with demanding customers by providing an extension of service coverage. Economical evaluations of transocean routes and global networks are of interest for decision makers responsible for business strategies as well as for operations. It is crucial to achieve appropriate judgements about which liner routes are profitable and how to build better service networks so that the companies' brand could be attractive to, e.g., shippers, including exporters, importers and forwarding agents. In this paper we discuss the corresponding trade-offs as well as related decision support systems of relevant service providers and companies.

Keywords— liner, port, logistics, networks, decision support system.

1. Introduction

Stemming from the inherent characteristics of world trade, the international shipping industry faces general issues of globalization, volatility, capital-intensity and periodicity. These issues, among others, provide the maritime shipping and logistics industry with a wealth of opportunities, however, with considerable uncertainties. The fact that liner carriers have already broken the barriers from being transportation providers towards being logistics service providers during the last two decades lets researchers consider related improvements and optimization after those extensions of business processes. Most of the top 20 liner carriers set up spin-offs providing different services, from local booking services up to 3rd party logistics services, combining the business advantages of tight linkages with the liners together with the negotiation freedom with demanding customers. A trade-off results from the relatively ambitious goals of the liner carriers and the marketing pressure of their spin-offs focusing on 3rd party logistics services, namely, the economical comparison between transocean routes and global networks. On one hand, as top liners deploy mega ships, economies of scale and single voyage efficiency are needed in order to accomplish the aim of unit cost saving. Thus, there is a need to focus on (long haul) transocean routes. On the other hand, for a 3rd party logistics provider (3PL), a global network with reliability and agility is crucial, too. Decision makers handling business strategies as well as operations, who are willing to resolve this trade-off, must be aware of which routes are productive and how to build up better service networks so that the brand and the reputation of affiliated companies is attractive to shippers, including exporters, importers and forwarding agents.

In this paper, besides presenting a literature review on a variety of papers relevant to the topic, we also attempt to discuss the evaluation and analysis of route choice and the optimization of networks. We start with the ingredients of related networks – ports and routes – and later extend by addressing different functions of the liners and the 3PL, as well as illustrating several criteria suitable for the selection of a transhipment hub as well as inland feeders. Regarding the dynamic competition and cooperation within the liner market, we finally sketch a theoretical framework, which may be of use regarding the development of decision support systems (DSS) of the liner companies, on how to build efficient service networks.

2. Ingredients of the networks – routes and ports

Logistics services could be identified as appropriate extensions of existing networks. Therefore, cf. [3], we interpret the connections among routes and ports of call as subnetworks. As the definition of logistics can be quite broad we need to focus. That is, in this paper we are mainly concerned with door-door service derived from long-distance shipping services. Short-distance inland distribution logistics without any shipping is not covered in this paper. In order to gain better understanding about the networks of shipping and logistics, the routes and the ports could be defined as the links and the nodes, respectively, as basic ingredients forming the networks.

2.1. Routes – links/ports – nodes

There are three main transocean lanes, namely transatlantic, transpacific and far-east to Europe [5], playing significant roles as the cheapest transportation mode serving commodity flows. Besides these main lanes, each liner carrier would arrange its services based on given freight requirements, thus, the required routings from the shippers motivate the liner carriers to construct complete world-wide networks. The lanes and routes connecting the ports can be viewed as the links in a graph with the seaborne transportation demands of the links denoted as weights.

Various researchers have addressed the definition of a "port". For instance, Carbone and De Martino [4] define: "*Ports have been natural sites for transhipment in order to transfer goods from one mode of transport to another. They have historically provided the link between maritime and inland transport, and the interface between the sea, rivers, roads and railways*". This definition is not fully comprehensive, as ports often also function as nodes without any mode diversity, e.g., from ship to ship. At least this statement indicates that ports are nodes with cargo in-flow and out-flow. (See, e.g., [31, 32] for a container terminal oriented survey.)

Considering the ports as nodes, we note that the liner carriers and their logistics spin-offs discussed in this paper are also of other characteristics – direct service networks together with indirect physical networks. For instance, a shipper as customer may book a door-to-door service in terms of local booking service[1] provided by a 3PL spin-off of the liner carriers. That is, the shipper and the service provider have direct service connections. However, it does not mean that this cargo freight is transported by the 3PL related liner carrier[2] only, not even within multi-modal transportation including inland-haul and short-sea distribution, if any, according to the service contract between the shipper and the 3PL. Therefore, indirect physical networks exist, which urges smooth and seamless connection.

2.2. Similarities among networks

Not only maritime shipping (transportation) networks have the features discussed in Subsection 2.1. Also some other (service) industries share the features which can be investigated by applying similar methods. It deserves to be noted that shipping and the logistics industry, as well as the telecommunications industry, share common characteristics of networks such as facility indivisibilities, technology interconnectedness and utility externalities, etc. A simplified comparison between the telecommunications and the port operations as well as freight shipping can be shown as follows. Similarities mainly exist regarding four aspects: generation and infrastructure development, distribution, mode choice and assignment. A brief comparison is that telecommunication service carries packages which contain data and messages; port operation moves containers either vertically up/off to/from ship or horizontally connected with trucks or trains; and also freight transport carries commodities from origins to destinations. Moreover, we can state that in all these "systems" we are concerned with consolidation and transhipment points.

Regarding the basic features of those industries and their similarities, theoretically, the literatures and research outcomes from each area could be applied to each other if done in an appropriate manner. For a review on the service network design for freight transportation see, e.g., Wieberneit [36] who specifically investigates tactical planning problems in freight transportation. Regarding the classification of the planning of a transport system, we refer to, e.g., [6]. In Section 3, we focus on the freight shipping industry.

[1]For reason of domestic maritime regulation and territory security, most countries do not authorise the foreign liners full authority of direct-booking. This is one of the reasons that foreign liners set up 3PL spin-offs as interfaces providing local-booking services to the shippers.

[2]It refers to a liner carrier who sets up a 3PL spin-off. Later, those two companies might become sister companies belonging to the same group.

3. Selection and preference of ports and networks

From a historical point of view, the main routes that contain lots of cargo desires are those routes firstly developed by ancient traders, and those nowadays need to be deployed with mega ships. However, taking basic logistics requirements into account, a superficial contradiction seems to arise from the liner companies and their 3PL spin-offs.

3.1. A superficial contradiction

On one hand, the target market niche of liner services is to provide transportation by visiting fixed ports according to pre-announced fixed schedules, meanwhile at a relatively stable freight of all kinds (FAK) price. More specifically, even the names of liners' vessels are settled and announced in advance once the liners are willing to provide liner services, and those container vessels are supposed to visit selected ports one by one in a timely fashion, also based on pre-announced fixed schedules. Cargo fitting into containers are shipped at settled prices (in this paper we ignore the issue of setting booking prices and the strategic contractual wholesale prices) disregarding what the cargoes really are. As a result, we refer to any TEU (twenty-foot equivalent unit; measurement of containers) as a profitable "unit".

A fundamentally common aim of the liner companies is to achieve economies of scale together with significant cost savings per unit, achieved by deploying bigger ships along profitable routes consisting of productive ports with deep drafts.

On the other hand, attractive service offerings provided by the logistics companies could be increased frequency, less quantity per shipment and higher agility based on customers' specialized requirements. Logistics companies providing 3rd party services with local booking authorities, especially those spin-offs of the liner companies considered here, are actually blooming since the last decline of the liner industry under the hope of attracting more customers from competitors providing similar liner services. Those 3PL spin-offs are endowed with the advantages of getting allocated capacities at lower contractual prices with their head companies or sister companies. Nevertheless, they are trying every effort to accomplish and fulfil door-to-door and even value-added services as well as to expand networks by means of visiting feeder ports and setting up inland distribution centres.

Then a superficial contradiction occurs between the selection of transhipment hubs and the expansion of networks under the capital constraint and management constraint of the head-corporation of the liner company and the involved 3PL. In this paper, we consider the network design problem as a strategic issue.

Note that we regard a spin-off of the liner carriers providing logistics service as 3PL. However, other researchers might rate liner carriers themselves as 3PL considering buyer and seller of the respective trade contract ([28], p. 252). Here we somewhat ignore the debate of who can actually be regarded as 3PL or even "4PL". Instead, we focus on the performance and value of the service networks. For a framework for evaluating 3PL see, e.g., [33].

3.2. Possible solutions to solve the contradiction

In this section, we investigate liner carriers and their 3PL spin-offs from a network theory perspective, which might shed some light on resolving the above mentioned contradiction. Applying network theory allows the liner carriers to optimize their current networks as well as aggregate potential partners' network [3]. Consequently, a multi-criteria optimization system should be set so that a rational selection on transhipment hubs and feeder ports could be accomplished. For a comprehensive literature review up to 2000 on freight transportation structuring from the viewpoint of choice processes we refer to [18]. Here we further discuss some other criteria in terms of networks, information as well as the 3PL spin-offs.

In practice, the selection of ports of call, including transhipment hubs and feeder ports, could be viewed as selection processes for business partners, no matter whether it relates to vertical or horizontal partnerships. However, before they become business partners, port operators, to great extent, might be competitors within the same industry. That is, players belonging to the same region, neighbouring each other and sharing overlapping hinterland, form a competitive relationship (e.g., the so-called North-Range in Europe). As some literatures address, financial health, adequate physical facilities, intangible assets [1, 10] are crucial as contributing factors during preliminarily screening the potential ports of call. Further references regarding port selection can be found in, e.g., [17, 24].

3.2.1. Criteria of hubs/transhipment hubs

Distribution network. One difference between hubs and transhipment hubs is whether there exists an advanced distribution network to connect to the hinterland. If there is an advanced distribution network, the hub may not only act as a media to move cargoes from one ship to another (cf. the term crossdocking in slightly different context), but also between different transport modes, e.g., from ships either to trains or to trucks. However, for many transhipment hubs, like Hong Kong or Singapore, a high percentage of the whole throughput refers to ship-to-ship movements. Thus, in such a case the hinterland distribution network is not of utmost importance (compared to, e.g., Hamburg). Most important are the free-port regulation and a sophisticated handling system that make the B/L transaction and water-water transhipment convenient.

Information system. Congestion, either on the seaside or on the landside could enlarge the total time of a vessel in the port, which would actually imply increased operational costs for the liner carriers. However, congestion free access to a port or congestion within the port is usually not one of the (main) criteria for choosing the hubs. As a matter of fact, several hubs suffer congestion quite often. It seems most important whether there is an efficient and effective information system to support the daily operations within the port so that even if congestion happens, a constructive solution would be suggested by the information system quickly. Recent discussion in this respect refers to so-called port community systems (see, e.g., www.dakosy.de for some example).

The 3PL spin-offs. In the hubs that the liner carrier or its corporation chooses, usually a related 3PL spin-off is set up, too, to ensure the convenience of the service that they could provide to the customers as a package. Comparing the local forwarder agent located in other feeder ports, the 3PL spin-off has stronger linkage with the liner carrier and, in return, might get more allocated capacities as support.

3.2.2. Criteria of feeder ports

Local forwarder agent. In practice, the selection of feeder ports is usually combined with the selection of the local forwarder agents. In most cases, if one forwarder agent distinguishes himself by his performance in one port, then other ports covered by this forwarder agent's business are probably also selected by the liner carrier as feeder ports. One superficial reason could be that the forwarder agent has a long cooperation with the liner carrier and gets used to follow all the managerial habits of the liner carrier which satisfies the liner carrier's requirements and further brings the liner carrier more freight. Another reason is that this forwarder agent could to great extent support the freight and fill capacities of the liner carrier by utilising his own network and attract shippers located in the hinterland. Considering the transocean routes initially constructed by the liner company, we define the extended inland or short-sea network of the local forwarder agent as the sub-network.

This phenomenon indicates that potential feeder ports would be selected due to their contribution to the original networks in fashion of better sub-network connection and accessibility. It should also be noted that such expected contribution might not happen as soon as the alternative feeder ports are added into the network, they might play their roles step by step. Unfortunately, as time goes by, the freight flow may amplify itself and then the profit-driven liner carriers may set up their own spin-off or stock-holding companies there instead of cooperating with the former forwarder agents. Consequently, this feeder port may even have the chance to be upgraded as hub within the ports of call of this liner.

Besides the practical criteria mentioned above, Lirn *et al.* [15] apply the analytic hierarchy process (AHP) as a method for evaluation and selection of transhipment ports from a global perspective. In addition, other researchers propose multi-criteria optimization for partner selection issues, which could be regarded as the amendment and development of an AHP application, see [9, 10].

4. Network optimization for a dynamic liner market

In this section we discuss aspects of optimizing service networks regarding the dynamic liner shipping market by taking into account the capacity of other sub-networks with a whole networks perspective. General concerns of cost efficiency in container shipping can be found, e.g., in [30].

4.1. Dynamics as a characteristic of the liner shipping industry

In spite of the cooperation among the liner carriers and other players involved in the liner shipping industry, many observations disclose the fact that the liner shipping industry is full of dynamics, including membership diversity, partnership reshuffling, network restructuring, etc. Rimmer [27] provides a historical description on the membership diversity among the liner shipping alliances up to the mid nineties. A more recent exposition of cooperation, mergers and acquisitions within the liner shipping industry is given by Notteboom [25]. Furthermore, for an up-to-date review on the dynamics existing in this industry see [29].

In short, the membership of the shipping alliances can switch from partnership towards being competitors and vice versa. This not only results in fleet capacity changes but also leads to diversity between the services that the alliances can provide. In this case, the related liner carriers' behavior of changing membership can be interpreted as attempting to combine new sub-networks with other players, no matter whether the other players are carriers or local in-land haul service providers.

4.2. Flexibility as response

Due to the dynamic environments of the transportation industry, flexibility plays a vital role if relevant companies are willing to survive. Reasons for the importance of flexibility include network externalities, as pointed out by David [7]; benefits of users/producers of the services are depending on the presence of other users/producers. Robinson ([28], p. 248) states that "*shipping lines are in the business of delivering value to buyers and sellers – and of capturing value to ensure they remain in business*". Considering the dynamics of the liner market, we address the flexibility of the network as one of the competitive advantages to ensure that shipping lines remain competitive and survive in business.

Once liner carriers have to compete in context of flexibility, the selection and integration of sub-networks becomes vital. Min and Guo [22] investigate the location of hub-seaports in the global supply chain network from the point of view of cooperative game theory. They develop a cooperative strategy in order to support the liner carriers and the shippers to determine optimal locations for the hub-seaports. However, our approach is slightly different as we do not assume the liner carriers and the local sub-network providers having binding agreements among each other. To some extent, we deepen our research based on the non-cooperative assumption, which is more realistic in the real business. As discussed in Subsection 2.1, routes and ports would be regarded as the basic features of the service networks of the liner carriers. The following four aspects need to be taken into account: generation of seaborne transportation, distribution of the shipping requirements, modal split and assignment of the shipping volume.

4.2.1. Zoning

While discussing ports serving the container flows, related regions are actually divided within the overall transportation networks by means of zoning. Zoning is a process that combines similar nodes into different zones and separates them from each other. Such zoning process depends a lot on the objectives of the networks, the available data, budget and time constraints as well as the zones homogeneity. Furthermore, due to the limited knowledge of all the details of every node almost all information (or expectation) of the nodes could be integrated into the "zone" and lateron each zone is reduced to a point. Then, spatial dimensions of a zone diminish. For instance, once ports *A* and *B* are integrated into one zone, the spatial distance between *A* and *B* is not important any more. In contrast, whether *A* or *B* would act as the hub of this zone would be an important decision. Once *A* acts as the hub and *B* acts as the feeder port, the assignment of the inbound and outbound links to and from this zone is related to the network design, while the volume between *A* and *B* within this zone is related to the sub-network design. In other words, one of the ports, say *A*, is selected as the central port because of advantageous transportation conditions while utilizing other ports, say *B*, as subsidiary within the zone. We note in passing, that a comprehensive survey of operations research approaches for the design of hub and spoke systems is provided in [34]. A simple adoption of hub and spoke systems to ship assignment is provided by Mourao and Pato [23].

4.2.2. Coding

The whole network is simplified by means of zoning and coding. Coding is a process that captures network links and centroids to represent the characteristics of the zone, respectively, instead of the former random links and nodes. That is, the network links and centroid are more relevant for the networks rather than the sub-network.

Now we are prepared to explain the behavior of the liner carriers: sometimes they set up 3PL spin-offs located in different areas and sometimes they simply select some local agents to act as the forwarder service and logistics service provider. However, the in-depth ideas are similar. The liner carriers set up their own 3PL spin-offs after zoning their current and potential traffic network and let the 3PL spin-off represent (the features of) this zone so that it could serve in the best possible way. As a different option, they select the local agents acting as a representative of a sub-network, whoever could contribute best to the whole network. Thus, the competition of the liner carriers, to some extent, is a competition of network integration. Furthermore, the turbulence of the liner market requires the flexibility of the network to ensure the just-in-time change. Once the circumstances or factors change as, e.g., observed with respect to the Panama Canal expansion, the sub-network and the whole network of the liner carriers should change accordingly to match customer demands and the circumstances.

In general, all the cargoes currently are served by the networks and the containerized shipments are transported from door to door, and during this procedure, at least two hubs are chosen (maybe more than two if the transhipment is included regarding the long distance). One hub locates in the zone of the origin, and the other hub locates in the zone of destination. Thus, the whole logistics procedure could be decomposed into the liner carrier's network and its 3PL spin-off/local service's sub-network. Furthermore, the sub-network selection and their connectivity are of great importance. We note that this is closely related to intermodal transportation problems, airline transportation networks as well as problems in telecommunications network design. For the latter see, e.g., the formal modeling approaches in [13, 20]. Route design in a specific liner shipping problem is considered in, e.g., [12].

In the following, we describe the problem from two aspects: sub-network selection and shipment distribution. Let $G = (N,A)$ be a graph consisting of a set of nodes N and a set of arcs A. G represents a physical network provided by the liner carriers and the logistics providers. Let K define a set of cargo shipments. A specific sea cargo shipment $k \in K$ is defined by an origin-destination or O−D pair, with $o(k)$ as origin and $d(k)$ as destination. The set of all paths from $o(k)$ to $d(k)$ for k is defined as P^k and the set of all O−D-pairs throughout the network is defined as P. The demand of a network or sub-network related to shipment k is denoted as d^k, which has to be transported from $o(k)$ to $d(k)$. In this paper important constraints such as time window constraints are not considered as key constraints as we ignore operational details. The main constraints refer to arc-capacity so that they could fulfil the demands d^k. Considering the integration of some networks, the total capacity of the involved links should be enough to cover the total demands. One might include binary decision variables indicating whether a sub-network is to be added to the whole network, or not. Another variable x_p^k is a nonnegative shipment flow variable, which indicates the flow of shipment $k \in K$ transported via the path $p \in P$, i.e., the amount of cargo to be shipped. F_p^k denotes the freight rate of the shipment $k \in K$ via the path $p \in P$. F_P^s denotes the freight rate of the additional shipment $s \in S$, which is attracted by the newly-added sub-network defined as S.

By adding appropriate arc inclusion indicator variables as well as flow variables we can model a multicommodity flow problem similar to those in telecommunications network design, see, e.g., [13, 20]. Here we concentrate on the objective function.

Let C_p^k be the shipment flow cost or variable cost of handling the goods per unit flow of k along path $p \in P^k$. The fixed costs of the network are not considered, because they are sunk costs in this problem. When the liner carrier decides to integrate with some potential sub-networks, the fixed cost of the carriers' network had already been invested before, and the amount of it would not be taken into account for the next stage. In contrast, the fixed costs of the potential sub-networks should be considered because they are among the main factors of the decision making process.

The objective function can then be formulated as follows:

$$\max \sum_{k \in K} \sum_{p \in P^k} F_p^k x_p^k + \sum_{s \in S} \sum_{p \in P^k} F_p^s x_p^s - \sum_{k \in K} \sum_{p \in P^k} C_p^k x_p^k ,$$

where:

$$x_p^k \geq 0, \quad \forall k \in K, \quad p \in P^k .$$

The objective is to maximize the total profit of the integrated network by taking into account not only the original shipping demand but also the additional shipping demand attracted by the improved network.

The nodes can be denoted as n_i and the zones can be denoted as z_i after zoning. Suppose that the liner carrier attempts to construct the global network or just to improve some part of the whole network. Figure 1 demonstrates the nodes, zones and the links of the sub-network and the whole network, respectively. We could not clearly separate the procedures of selection (set up 3PL spin-offs or select local agents) and zoning because they actually happen almost at the same time. However, slight differences still exist. As for setting up a 3PL spin-off of the liner carrier, it might happen after zoning because at the moment of location selection the liner carrier has already build up a global service network and most probably the headquarter of the 3PL spin-off will be located just in the centroid of the zone. In contrast, the selection of the local agents may influence the zoning of the liner carrier because some of the local agents are so strong that the shipping volume of the related zone changes too much. However, the common idea of the setting up and selection is whether such decision would contribute to the payoff of the whole network as well as the sub-network itself. Suppose that one liner carrier attempts to cover the two main lands. The zoning process follows the criteria of covering as many freight nodes and simplifying the whole area as much as possible. The coding process lets the z_i represent instead of n_i, which tremendously decrease the links and the voyage time of the vessels. However, once comparing the potential options of the local agents of z_6 and z_7, an overlap of these two zones is found. This infers that if the first local agent of z_6 is not strong, the initial zoning result can be obtained, but if another alternative local agent is to be integrated, then the zoning of z_6 and z_7 shall be reorganized.

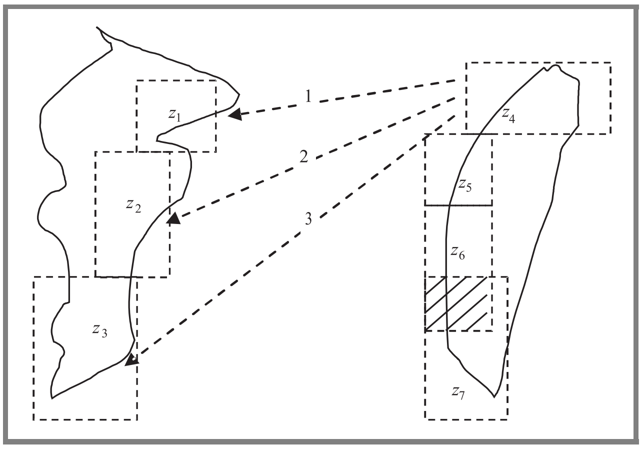


Fig. 1. The zoning of the origin and the destination of long distance transportation.

In Fig. 1, without loss of generality, we take z_4 as the origin area and the left side as the options of destination, including z_1, z_2 and z_3. The optimal route from the right hand side to the left hand side of this figure depends on the sub-networks inside the zones z_1, z_2 and z_3 and their connectivity.

Regarding the integration of any sub-network into the whole network, both the payoff of the sub-network provider, in this case, a 3PL or local agent, and the payoff of the whole network must be positive and bigger than the former stages. Otherwise, such integration usually makes no sense to the liner carriers.

For possible heuristics to solve and validate some concept proposed in this paper, we refer to [35]. Regarding a mathematical proof of a similar port-of-call scheduling problem we refer to [19]. In addition, an interesting case study including six European ports in the context of port selection in the hinterland of Europe is [8].

5. Decision support systems in transportation companies

As indicated in Section 3, competition among the liner carriers currently relies on the implementation of the service networks by means of selecting sub-network providers and cooperating with them. Furthermore, in order to have a smooth coordination and integration of different sub-networks some sophisticated information systems are necessary.

5.1. Liner carriers and port operators integration for efficient supply chain management

Stepping back in history and the development of trade, transportation and logistics, liner carriers or shipping carriers in general were pure traders centuries ago. However, nowadays they tend to have the ambition of being more comprehensive players. The shipping carriers attempt to touch inland-haul service, short-sea connections and certainly logistics service mainly related to door-door transportation (here we do not refer to the so-called value-added activities inside manufacturing factories which are always included as logistics, too).

Shippers and consignees are exporting and importing the cargoes and pay the freight rates, accordingly. However, as they only have direct service contracts with the liner carriers or the 3PL rather than with the port operators [29], the detailed operations between logistics providers and port operators are of less interest for them. Consequently, the efficient and effective integrated services including transportation services and port operations would be most welcomed by the customers, i.e., the shippers and the consignees.

In order to obtain better performance of service integration, DSS are of great importance for the liner carriers. Regarding the difference between DSS and decision making systems we distinguish whether the systems recommend several potential actions or automatically implement actions [16]. For the current solution methodologies, optimization-based solutions of information systems focus more on the average demands and requirements under static conditions, and simulation-based solutions accommodate the system dynamics which could be more suitable for the real-world business [16]. Furthermore, heuristic-based models contain the capability taking into account almost all network configurations providing optimized solutions accordingly.

5.2. Decision support system applications in liner and logistics companies

Since the last decade internet-based business (or e-business) activities have become a new technological challenge for the shipping industry. However, beyond the introduction of electronic data interchange (EDI) little systematic and theoretical research on e-business has been undertaken within this area so far. Therefore, we attempt to investigate the application of information systems in the shipping industry (the container shipping industry is focused in this paper) and their impacts of e-business on the container shipping industry in order to provide the liner carriers with the managerial recommendations accordingly. For a literature review on general business dynamics and the technology strategies of six different e-business models in the container shipping industry see [2]. Moreover, we should note that various areas lack the practical application of DSS, largely due to the lack of unified generally applicable systems, cf. [11].

While business activities could be divided into operational, tactical and strategic activities, the respective sub-information systems perform various functions. End users of a container shipping company could be basically distinguished regarding activities along those time horizons from strategy promoters up to in-putters of daily operational data. These include vessel positions, container status, service requirements, payment transactions and so on.

As an example we consider Maersk Sealand which is regarded as a benchmark from almost every aspect in the shipping industry. We focus on a "handwaving" description of its information systems applications as well as implementations. The whole information system could be called MGM, consisting of three subsystems, namely MARS, GCSS and FACT, aiming at handling contracts, booking accomplishment and finance accounting, respectively.

Once we start our observation from the most basic activities – slot booking and bill of lading (B/L) issuing – of a container shipping company, it would shed light on the whole applied information system. As shown in Fig. 2, the exporter books capacity based on his planned cargo transportation[3], which is going to be transported to the importer. To simplify the process, we regard the exporter as the shipper and the importer as the consignee regardless the pure medium trader who actually does not produce or own cargo. The pure medium trader gains profits by buying and selling cargo at different price, or maybe only transacting the B/L rather than cargo itself.

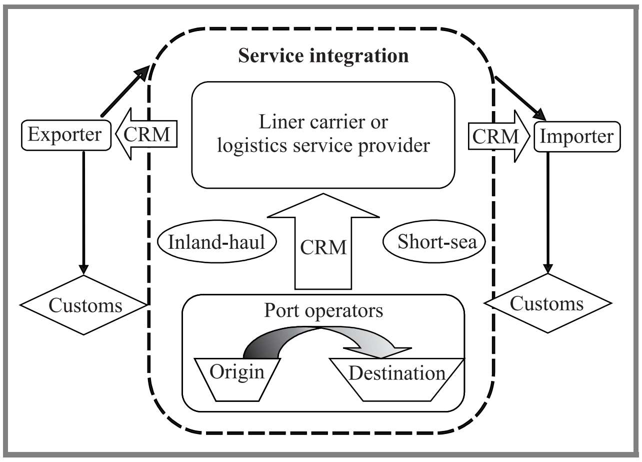


Fig. 2. The service integration and its cash/cargo/information flows (CRM – customer relationship management).

There are older and mature information systems applied within Maersk Sealand, namely MARS and RKDS, which help sales representatives and customer service staff to advise transportation services, arrange routes, input and output data. The interfaces of those systems were long-time criticized as not being user-friendly enough. They are still simultaneously applied together with a new system called global customer service system (GCSS) designed and developed by IBM. The GCSS is currently used mainly by the customer service group globally providing functions like routing, tracking, on-line publishing, etc. A "rater" – providing customer service – is supposed to use GCSS to figure out the service contract of a shipper and fix the service price according to this shipper's booking. Beyond expectation, more "raters" are now hired in Maersk Sealand than during the period of applying the old systems, as it turns out to be even harder to exchange data using the new information system. Moreover, the interface of GCSS with other subsystems is not as smooth as expected. Manual work is arranged to supplement system problems.

Regarding tactic and strategic level business management, profit judgement and risk evaluation would be two main aspects for which information systems perform decision support functionality.

Process standardization. From a customers' perspective, requirements would be well satisfied if they are met timely and specifically. In the past, once the requests of VIP customers change, the workflow of the carriers may change as well. However, implementing a new information system results in a situation where most customers are regarded exactly the same no matter what amount of cargo they transport, while those customized requirements would be noted in the specific entries inside the systems. Due to standardized workflows within the system even exception handling is assumed to be more streamlined especially when faced by employees who are new to specific situations.

Profit pre-analysis. Continuous deficits push decision makers to consider whether persisting transportation operations along the involved routes and ports are profitable or not. Various aspects are vital since any change of the liner routes and logistics networks would lead amounts of investment not only in marketing surveys but also in acquisition of infrastructures including vessels and cranes, etc. That is, it is a capital-condensed and cost-sensitive industry.

Similar to the operational systems of other liner carriers and logistics companies, MARS in Maersk Sealand provides distinctive options for cost per unit and expected benefit calculation for various types of containers regarding, e.g., volumes they occupy on deck and in haul (such as 20DC/40DC, i.e., a single 20- or 40-foot container containing dry cargo; HC, i.e. a 45-foot high cube container; etc.). Currently, an SAP R/3 package is implemented, namely financial accounting for container transport (FACT), and it is planned to be released by the end of 2008.

Risk evaluation. Risk evaluation based on historical data, service simulation, and expert judgement is of importance to demonstrate whether to accept specific transportation requirements. For risk management considerations regarding other types of cargo, such as crude and product oil see, e.g., [21].

[3]Here we assume trade contracts in terms of CIF (cost, insurance and freight). In applying other INCOTERMS, the analysis is similar. For basic knowledge on INCOTERMS, we refer to, e.g., [14].

Different types of containers as well as cargo need to be handled differently, especially reefer and hazmat containers. On December 18, 2006, the REACH (registration, evaluation, authorisation and restriction of chemicals) regulation was formally adopted by the European Union and is enforced since June 1, 2007. In order to save the testing cost on chemicals and to get an overview about which studies are available, a system which could serve as data sharing platform is currently under construction. A supplementary but vital requirement of this system is to ensure that not only manufacturers and importers but also their customers and distributors have the information they need to use and transport chemicals safely. Information relating to health, safety and environment properties, and risk measurement is required to be shared along the supply chain. Commercially sensitive information is not required to be exchanged [26]. Although REACH has just been put into force recently, its effect on information flow management within supply chains is regarded as huge.

However, it should be noted that, due to fast EDI processes, the information centre of the liner carrier need not be the centroid of any zone defined in Section 4. Actually, some of the information centres of the liner carriers are even located far from hubs, following various criteria such as human resources availability and cost. The geographic location of an information centre is not a key issue in this paper and it may be viewed as a fictitious node that contributes to the whole service network.

6. Conclusions and further research

In this paper, we have discussed the network structure of the maritime liner shipping companies and their spin-offs providing 3rd party logistics services. Commonalities with intermodal transportation in general as well as with telecommunications network design may serve as a means for advancing the subject. Moreover, game theoretical approaches may help to support strategic as well as tactical decision making in liner shipping. This may involve the assumption of cooperative as well as non-cooperative behaviors of involved players on different levels.

References

[1] K. W. Artz and T. H. Brush, "Asset specificity, uncertainty and relational norms: an examination of coordination costs in collaborative strategic alliances", *J. Econom. Behav. Organiz.*, vol. 41, no. 4, pp. 337–362, 2000.

[2] J. H. Auh, "Analysis of the impacts of internet-based business activities on the container shipping industry: the system dynamics modelling approach with the framework of technological evolution", Ph.D. thesis, Cambridge, MIT, 2005.

[3] A. S. Bergantino and A. W. Veenstra, "Interconnection and coordination: an application of network theory to liner shipping", *Int. J. Marit. Econom.*, vol. 4, no. 3, pp. 231–248, 2002.

[4] V. Carbone and M. De Martino, "The changing role of ports in supply-chain-management: an empirical analysis", *Marit. Policy Manage.*, vol. 30, no. 4, pp. 305–320, 2003.

[5] P. Cariou, "Strategic alliances in liner shipping: an analysis of "operation synergies", Ph.D. thesis, University of Nantes, 2000.

[6] T. G. Crainic, "Service network design in freight transportation planning", *Eur. J. Oper. Res.*, vol. 122, pp. 272–288, 2000.

[7] P. A. David, "Information network economics: externalities, innovations and evolution", in *The Economics of Information Networks*, C. Antonelli, Ed. Amsterdam: North Holland, 1992, pp. 103–106.

[8] P. W. de Langen, "Port competition and selection in contestable hinterlands; the case of Austria", *Eur. J. Transp. Infrastr. Res.*, vol. 7, no. 1, pp. 1–14, 2007.

[9] J.-F. Ding and G.-S. Liang,"Using fuzzy MCDM to select partners of strategic alliances for liner shipping", *Inform. Sci.*, vol. 173, pp. 197–225, 2005.

[10] J.-F. Ding, G.-S. Liang, C.-H. Yeh, and Y.-C. Yeh, "A fuzzy multi-criteria decision-making model for the selection of courier service providers: an empirical study from shippers' perspective in Taiwan", *Marit. Econom. Logist.*, vol. 7, pp. 250–261, 2005.

[11] J. Engelke and H. Simon, "Decision support systems in marketing", *J. Bus. Mark. Manage.*, vol. 1, pp. 289–307, 2007.

[12] K. Fagerholt, "Designing optimal routes in a liner shipping problem", *Marit. Policy Manage.*, vol. 31, no. 4, pp. 259–268, 2004.

[13] H. Höller and S. Voß, "Software tools for a multilayer network design", *J. Telecommun. Inform. Technol.*, no. 3, pp. 57–63, 2005.

[14] INCOTERMS, 2007 (last call 15.12.2007), http://www.iccwbo.org/incoterms/id3045/index.html

[15] T. C. Lirn, H. A. Thanopoulou, M. J. Beynon, and A. K. C. Beresford, "An application of AHP on transhipment port selection: a global perspective", *Marit. Econom. Logist.*, vol. 6, pp. 70–91, 2004.

[16] G. M. Magableh, "A dynamic replenishment system for integrating supply chain functions", *Marit. Econom. Logist.*, vol. 9, pp. 52–66, 2007.

[17] M. Malchow and A. Kanafani, "A disaggregate analysis of factors influencing port selection", *Marit. Policy Manage.*, vol. 28, no. 3, pp. 265–277, 2001.

[18] J. Mangan, C. Lalwani, and B. Gardner, "Identifying relevant variables and modelling the choice process in freight transportation", *Int. J. Marit. Econom.*, vol. 3, pp. 278–297, 2001.

[19] T. Matsumura and D. Shimizu, "Cournot and Bertrand in shipping models with circular markets", *Reg. Sci.*, vol. 85, no. 4, pp. 585–598, 2006.

[20] B. Melian, H. Höller, and S. Voß, "Designing WDM networks by a variable neighborhood search", *J. Telecommun. Inform. Technol.*, no. 4, pp. 15–20, 2006.

[21] J. R. W. Merrick, J. R. Van Dorp, J. Harrald, T. Mazzuchi, J. E. Spahn, and M. Grabowski, "A systems approach to managing oil transportation risk in prince William Sound", *Syst. Eng.*, vol. 3, no. 3, pp. 128–142. 2000.

[22] H. Min and Z. Guo, "The location of hub-seaports in the global supply chain network using a cooperative competition strategy", *Int. J. Integr. Supply Manage.*, vol. 1, no. 1, pp. 51–63, 2004.

[23] M. C. Mourao and M. V. Pato, "Ship assignment with hub and spoke constraints", *Marit. Policy Manage.*, vol. 29, no. 2, pp. 135–150, 2001.

[24] A.-S. Nir, K. Lin, and G.-S. Liang, "Port choice behaviour-from the perspective of the shipper", *Marit. Policy Manage.*, vol. 30, no. 2, pp. 165–173, 2003.

[25] T. E. Notteboom, "Container shipping and ports: an overview", *Rev. Netw. Econom.*, vol. 3, no. 2, pp. 86–106, 2004.

[26] REACH, 2007 (access date Apr. 24th, 2007), http://ec.europa.eu/environment/chemicals/reach/reach_in_brief04_09_15.pdf

[27] P. J. Rimmer, "Ocean liner shipping services: corporate restructuring and port selection/competition", *Asia Pac. Viewp.*, vol. 39, no. 2, pp. 193–208, 1998.

[28] R. Robinson, "Liner shipping strategy, network structuring and competitive advantage: a chain systems perspective", *Res. Transp. Econom.*, vol. 12, pp. 247–289, 2005.

[29] X. Shi and S. Voß, "Container terminal operations under the influence of shipping alliances", in *Risk Management in Port Operations, Logistics and Supply Chain Security*, K. Bichou, M. G. H. Bell, and A. Evans, Eds. London: Informa, 2007, pp. 135–163.

[30] D. P. Song, J. Zhang, J. Carter, T. Field, J. Marshall, J. Polak, K. Schumacher, P. Sinha-Ray, and J. Woods, "On cost-efficiency of the global container shipping network", *Marit. Policy Manage.*, vol. 32, no. 1, pp. 15–30, 2005.

[31] R. Stahlbock and S. Voß, "Operations research at container terminals – a literature update", *OR Spectrum*, vol. 30, pp. 1–52, 2008.

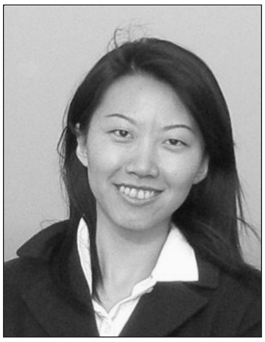
[32] D. Steenken, S. Voß, and R. Stahlbock, "Container terminal operation and operations research – a classification and literature review", *OR Spectrum*, vol. 26, pp. 3–49, 2004.

[33] G. Vaidyanathan, "A framework for evaluating third-party logistics", *Commun. ACM*, vol. 48 no. 1, pp. 89–94, 2005.

[34] B. Wagner, *Hub&Spoke-Netzwerke in der Logistik: Modellbasierte Lösungsansätze für ihr Design*. Wiesbaden: Gabler, 2006.

[35] K. Weng, C. Yang, and Y. Ma, *Two Artificial Intelligence Heuristics in Solving Multiple Allocation Hub Maximal Covering Problem*, *LNCS*. Berlin-Heidelberg: Springer, 2006, vol. 4113, pp. 737–744.

[36] N. Wieberneit, "Service network design for freight transportation: a review", *OR Spectrum*, vol. 30, pp. 77–112, 2008.



Xiaoning Shi is a Ph.D. candidate jointly affiliated with the Institute of Information Systems at the University of Hamburg, Germany, and the Department of Management at the Shanghai Jiao Tong University, P. R. China. Her position in Hamburg is supported by the DAAD (Deutscher Akademischer Austauschdienst) and CSC (China Scholarship Council). Formerly, she was working as a teaching assistant and later as a lecturer at the Institute of International Shipping, Department of Navel Architecture Ocean and Civil Engineering, Shanghai Jiao Tong University. In 2005, she was a visiting research assistant at the University of Hong Kong and in 2002 she was working at Maersk-Sealand Shipping Corporation, Shanghai. She holds Bachelor and Master degrees in engineering from Dalian Maritime University (China). Her current research interests are in econometrical and game theoretical applications in the shipping/port industry and optimization of logistics networks.
e-mail: shi@econ.uni-hamburg.de
Institute of Information Systems
University of Hamburg
Von-Melle-Park 5
D-20146 Hamburg, Germany

Shanghai Jiao Tong University
1954 Hua Shan Road
Shanghai 200030, P. R. China

Stefan Voß is Professor and Director of the Institute of Information Systems at the University of Hamburg, Germany. Previous positions include full Professor and Head of the Department of Business Administration, Information Systems and Information Management at the University of Technology Braunschweig, Germany, from 1995 up to 2002. He holds degrees in mathematics (diploma) and economics from the University of Hamburg and a Ph.D. and the habilitation from the University of Technology Darmstadt. His current research interests are in quantitative/information systems approaches to supply chain management and logistics including public mass transit and telecommunications. He is author and co-author of several books and numerous papers in various journals. Professor Voß serves on the editorial board of some journals including being Editor of "Netnomics", Editor of "Annals of Information Systems", Associate Editor of "INFORMS Journal on Computing" and Area Editor of "Journal of Heuristics". He is frequently organizing workshops and conferences. Furthermore, he is consulting with several companies.
e-mail: stefan.voss@uni-hamburg.de
Institute of Information Systems
University of Hamburg
Von-Melle-Park 5
D-20146 Hamburg, Germany

Paper

Negotiations on regulated markets

Sylwester Laskowski

Abstract—The paper considers some problems of negotiations between competitive subjects on regulated market. It is assumed that two subjects (players) have to compete with each other on the retail market and cooperate on the wholesale market. The wholesale market is regulated. The role of the regulator is to support players in negotiations, especially by introducing recommended solutions when the negotiations were broken off. It is considered how introducing a recommended solutions influence the process of negotiations on the wholesale market and a decision problem of choosing retail strategy, that precedes the process of negotiations. A decision problems of a regulator are also formulated. The problems are discussed in context of competition and cooperations between operators on the telecommunications services market.

Keywords— negotiations, regulations, market games.

1. Introduction

Competition and cooperation, these two radically different kinds of co-existence meet at one time and one place during a negotiation process. If both (or every of) sides are convinced, that the best way to realize their own aims is to cooperate with the other side, then they usually meet both together at a negotiations table, and try to find solution that would be better for both of them, then the solutions accessible outside the table. Limited resources, however, involves them into competition: each side (the most frequently) wants to get the highest or the best part of the divided "cake". So competition is natural part of the negotiations process. However it is specific kind of competition. This competition arises only, because both parts want or have to cooperate. So in negotiations competition is something like a daughter of the cooperation.

In networking businesses like, e.g., telecommunications there exist also opposite case: the players are sometimes forced to cooperate, because they operates on the same market (what makes them competitors) and so they have to interconnect their networks to provide full services for their own customers. So in such cases cooperation is a daughter of competition. However sometimes this is a daughter by one part unwanted. The strongest part, the higher faith that competition is a sufficient (or the best) tool for obtaining intended goals, and the lowest will for cooperation. So increasing the power outside the negotiations table increases also the power at the negotiations table, and at the extreme case the powerful player does not want to negotiate at all. Paradoxically this can also eliminate competition, because without interconnection weaker player cannot operates profitably. This is the reason, why in some cases it is impossible to transform a monopolistic market into competitive market (or safe from the opposite process), without active support by the third party.

On the telecommunications services market the role of such third party plays national regulator authority (regulator). On of the main instruments for supporting competition, which can be used by the regulator, is the possibility of forcing the strongest side to negotiate, and if the negotiations were broken off, possibility of introduction recommended solutions for forced cooperation.

The paper considers some problems of negotiations between competitive operators on regulated market. It is assumed that two players – A and B have to compete with each other on the retail market (in relation to the end customers) and cooperate on the wholesale market (interconnection). The wholesale market is regulated. The role of the regulator is to support players in negotiations, especially by introducing recommended solutions if the negotiations were broken off.

It is considered how introducing a recommended solutions influence the process of negotiations on the wholesale market and a decision problem of choosing retail strategy, that precedes the process of negotiations. A decision problems of a regulator are also formulated.

2. Definition of the negotiation power

The negotiation (bargaining) power can be defined into two ways:

- as a positive power, that enables the player to obtain a good outcome for himself;
- as a negative (antagonistic) power, that enables the player to deteriorate the outcome of the other player.

Let's denote the positive power of – respectively, player A and B – by α_p^A and α_p^B. The negative powers will be denote by α_n^A and α_n^B. The highest negative power of the player (e.g., A) the lowest positive power of the other player (B) and vice versa. Assuming that the bargaining powers sums up to one we have:

$$\alpha_p^A = 1 - \alpha_n^B, \tag{1}$$

$$\alpha_p^B = 1 - \alpha_n^A. \tag{2}$$

We assume that the negotiation power (positive and so negative) comes from the best alternative to a negotiated agreement (BATNA) [4, 11, 14] – a solution, that the player can obtain if the negotiations were broken of. So braking the negotiations is one of the possible solutions of the negotiation process. However this solution is not a unique one, because outside the table players can play as positively so negatively. Hence a situation of breaking the negotiations can be described by for different strategies of playing outside the table:

h_{pp} – both players play in a positive way;

h_{nn} – both players play in a negative way;

h_{pn} – player A plays positively, player B plays negatively;

h_{np} – player A plays negatively, player B plays positively.

We assume that if the player played in positive way then (independently of the way of playing by the other player) he/she would get higher value of the payoff function than if he/she played negatively. We also assume that he/she would get higher value of the payoff function if the other player played positively than if he/she played negatively. Hence we get:

$$V^A(h_{pp}) \geq V^A(h_{pn}) \geq V^A(h_{np}) \geq V^A(h_{nn}), \tag{3}$$

$$V^B(h_{pp}) \geq V^B(h_{np}) \geq V^B(h_{pn}) \geq V^B(h_{nn}). \tag{4}$$

Which strategy, and so which value of the payoff function defines BATNA of the players? BATNA is the best alternative so it should reflects the positive playing by the player. But it also must be attainable (independently of the way of playing by the other player). So $V(h_{pp})$ cannot define BATNA of the players because it is not attainable if one of the players played in a negative way. So the answer is h_{pn} for player A and h_{np} for player B. These strategies leads to the highest payoffs from these that players can be sure to obtain.

Using the concept of BATNA we can define the negotiation powers of the players as follows:

$$\alpha_p^A = 1 - \alpha_n^B = \frac{V^A(h_{pn}) - V^{A\min}}{V^{A\max} - V^{A\min}}, \tag{5}$$

$$\alpha_p^B = 1 - \alpha_n^A = \frac{V^B(h_{np}) - V^{B\min}}{V^{B\max} - V^{B\min}}, \tag{6}$$

where:

$$V^{A\max} = \max_l V^A(h_l), \tag{7}$$

$$V^{B\max} = \max_l V^B(h_l), \tag{8}$$

$$V^{A\min} = \min_l V^A(h_l), \tag{9}$$

$$V^{B\min} = \min_l V^B(h_l). \tag{10}$$

For numerical reasons it can be sometimes useful to make small modification:

$$\alpha_{\varepsilon p}^A = 1 - \alpha_{\varepsilon n}^B = \frac{\max\left\{V^A(h_{pn}) - V^{A\min}, \varepsilon\right\}}{\max\left\{V^{A\max} - V^{A\min}, \varepsilon\right\}}, \tag{11}$$

$$\alpha_{\varepsilon p}^B = 1 - \alpha_{\varepsilon n}^A = \frac{\max\left\{V^B(h_{np}) - V^{B\min}, \varepsilon\right\}}{\max\left\{V^{B\max} - V^{B\min}, \varepsilon\right\}}, \tag{12}$$

where ε is a small value.

As we see from Eqs. (1) and (2) there is no direct relation between positive and negative negotiation power of a given player (e.g., between α_p^A and α_n^A). Also there is no direct relation between positive negotiation power of one player and positive negotiation power of the other player (between α_p^A and α_p^B). Similarly there is no direct relation between negative powers of the players (between α_n^A and α_n^B).

However if we compare and sum up sides of the Eqs. (1) and (2) then we get:

$$\alpha_p^A - \alpha_n^A = \alpha_p^B - \alpha_n^B, \tag{13}$$

$$\frac{\alpha_p^A + \alpha_n^A}{2} + \frac{\alpha_p^B + \alpha_n^B}{2} = 1. \tag{14}$$

Equation (14) can be expressed as

$$\alpha^A + \alpha^B = 1, \tag{15}$$

where α^A and α^B can be treated as aggregated negotiation powers of the players A and B and are:

$$\alpha^A = \frac{\alpha_p^A + \alpha_n^A}{2}, \tag{16}$$

$$\alpha^B = \frac{\alpha_p^B + \alpha_n^B}{2}. \tag{17}$$

3. The impact of recommended solution on the process of negotiations

In the case of telecommunications market, the operators must negotiate the rules of the interconnection agreements. However in many cases there is a high difference between negotiation powers of the players, especially when one of the sides is an incumbent operator. New entrant of the market has usually much smaller network, and so much less end users connected to its network, than operating for long time incumbent. So it is necessary for new operator to interconnect its network to incumbent‘s network. But it is not necessary for the incumbent. This makes that the incumbent has very strong and new entrant very weak BATNA in the negotiation process. This difference, without protection by the third side, can be exploited by the stronger player with large disadvantage of the weaker player.

For the reasons of promotion fair competition the role of a regulator is to support new entrants in the negotiation process. The main instrument for doing this is possibility of recommending reference solutions for the negotiated interconnection agreement, and in the case of breaking off the negotiation without any agreement, possibility of forcing this solutions.

Now we will examine how such a recommended solution, which we denote as h^*, influences the negotiation process.

3.1. Disclosure the values of BATNA

Recommended by a regulator strategy h^* defines new BATNA of the players. In the case without a recommended strategy h^* BATNA of the both players is defined by two different strategies:

h_{pn} – defines BATNA of the player A;

h_{np} – defines BATNA of the player B.

Now BATNAs of both players are defined by strategy h^*.

Probably the most important thing for the process of negotiation is not that h^* defines *new* BATNAs, but that these BATNAs *are commonly known*. BATNA defines the positive negotiations power of the players: a player does not agree on the strategy that gives him worse outcome than his BATNA. So players would like to have as high BATNA as possible. But they also want the other player to think that it is also hight if really it is not. So in many situations players misrepresents and lie one another on the true value of their BATNA. Recommendation of a strategy h^* defines new and commonly known BATNA, and so makes such misrepresentations and lies impossible.

3.2. Integration of BATNAs

In the case without a recommended strategy h^* player A could be sure, that he/she could get the payoff not smaller than $V^A(h_{pn})$, and the player B could be sure, that he/she could get the payoff not smaller than $V^B(h_{np})$. But in the case of existing recommended strategy h^*, which can be chosen by both players, player A can be sure, that he/she could get the payoff not smaller than $V^A(h^*)$, and player B can be sure, that he/she could get the payoff not smaller than $V^B(h^*)$. So now the BATNAs of both players is determined by the same strategy – h^*. So we have something what can be called as *integration* of BATNAs. In some cases this fact can be very helpful in the negotiations process.

Example 1. Let's consider a simple example of the negotiations between player A and B. There are two accessible strategies at the negotiations table: h_1 and h_2, and four strategies outside the negotiations table: h_{pp}, h_{nn}, h_{pn} and h_{np}. For each strategy players obtains different values of payoff function, as in Table 1, which are commonly known. For example if during a negotiations players chose strategy h_1, then player A would obtain $V^A(h_1) = 10$ and player B would obtain $V^B(h_1) = 6$.

Table 1
An example of the positive impact of the integration of BATNAs on the negotiations process

Strategy	$[V^A(.), V^B(.)]$
h_1	[10,6]
h_2	[6,10]
h_{pp}	[5,4]
h_{pn}	[2,1]
h_{np}	[1,3]
h_{nn}	[0,0]

In the case that there is not a recommended strategy h^* BATNA of the player A is determined by strategy h_{pn} and equals $V^A(h_{pn}) = 2$. In this case BATNA of the player B is determined by strategy h_{np} and equals $V^A(h_{np}) = 3$. Negotiations game has two effective solutions (both obtained at the table) for strategy h_1 and h_2. Player A prefers the solution for strategy h_1 because than he obtains $V^A(h_1) = 10$, but player B prefers the solution for strategy h_2, cause then he obtains $V^A(h_2) = 10$. So both players would like to choose different strategy, and different solution. What is important, both players have strong argumentation for choosing prefered by them solution. Player B can argue like this: „Strategy h_2 leads to the solution, for which proportion of outcomes $(6/10)$ is nearer to the proportion of BATNAs $(2/3)$ then for strategy h_1 $(10/6)$, so choosing strategy h_2 is fair solution." But the answer of player A can be also convincing: „I do not want to play negatively outside the negotiations table (what is assumed in the case of choosing strategy h_{np} outside the table). Why do You want to do so? Fair solution outside the table is $[5,4]$, when we both play positively. So h_1 leads to the solution, for which proportion of outcomes $(10/6)$ is the nearest to the proportion of outcomes for the fair solution outside the table $(5/4)$, and choosing strategy h_1 is really fair solution!"

Both parts have strong argumentation, and if any different (creative and profitable) solution would not be found, than negotiations can be broken off, and the result of the game would be inefficient. The problem arises from existing several different reference solutions outside the negotiations table: BATNA of the player A, BATNA of the player B, result for the case that both players play positively (strategy h_{pp}).

Recommendation of the strategy h^* gives new or makes stronger one of the before existing reference solution. In some cases it can exclude every different references. For example, if a regulator recommends strategy $h^* = h_{pp}$, than there would be only one important reference point for the negotiating party. This strategy would define new BATNAs of the players and so, any different strategy outside the table could not be chosen. So the player A could argue that the proportion of the payoffs for strategy $h^* = h_{pp}$ defines fair proportion of the outcomes, and so strategy h_1 is more fair than strategy h_2 at the table, and the player B has not comparatively strong argument for choosing strategy h_2. This can make negotiations simpler and faster.

Of course, recommendation of the strategy h_{np}, although also integrates BATNAs of the players, would have not so strong impact on the improving of the negotiations. If for example a regulator recommends strategy $h^* = h_{np}$, player A can still argue that he does not want to play negatively outside the table, and press for regarding h_{pp} as a fair reference point. Obviously this situation makes argumentation of player B a little stronger. In the case that there was not a recommended solution it was impossible to realize simultaneously the BATNA of the players, because it would require to choose two different strategies at the same time, what is impossible. So his/her argumentation on proportion of BATNAs was somehow weak. Now, when h^* defines new BATNA of both players this argumentation becomes stronger.

Obviously recommendation of the strategy h_{pn}, due to the corresponding values of the outcomes $[2,1]$ would make radically stronger argumentation of the player A. Now proportion of (new) BATNA $(2/1)$ and proportion of the outcomes for positive playing $(5/4)$ indicates on the strategy h_1 as a much more fair solution then h_2.

□

Above illustrated example shows that changing the recommended strategy h^* changes the power of argumentation for choosing different strategies at the negotiations table. In fact changing (and before it introducing) a recommended strategy h^* changes the negotiations power of the players. Now we will examine this subject with distinction on the positive and negative negotiations powers of the players.

3.3. The impact of recommended solutions on the negotiations power of the players

As it was sad above, recommended solution integrates BATNA of the players, so that strategies h_{np} and h_{pn} no longer defines it, but it is determined by strategy h^*. Each player cannot be sure that he/she can obtain the outcome higher then determined by strategy h^*, because always the other player can brake off the negotiations and require the regulator to introducing strategy h^*. But from the other side, each of them can be sure that he/she can obtain the outcome not smaller than determined by h^*. So h^* defined new BATNA of both players.

This integration of BATNAs can simplify negotiations process by reducing a reference solutions outside the negotiations table for fair one at the table. However, as it was illustrated in Example 1, this depends on which strategy is recommended. Recommending of some strategies can make negotiations simpler and faster, but recommending some different can make negotiations more difficult and slower. Now we will examine this problem, by analysing how different recommended strategies influence the negotiations power of the players, with distinction on positive and negative power.

The reasoning is simple: the higher BATNA of the player, the stronger his/her positive negotiations power, and the lower negative power of the other player. So we can write the following relations:

- If recommended solution is under BATNA of the player A: $V^A(h^*) < V^A(h_{pn})$ than decreases positive negotiations power of the player A and increases negative negotiations power of the player B.
- If recommended solution is above BATNA of the player A: $V^A(h^*) > V^A(h_{pn})$ than increases positive negotiations power of the player A and decreases negative negotiations power of the player B.
- If recommended solution is under BATNA of the player B: $V^B(h^*) < V^B(h_{np})$ than decreases positive negotiations power of the player B and increases negative negotiations power of the player A.
- If recommended solution is above BATNA of the player B: $V^B(h^*) > V^B(h_{np})$ than increases positive negotiations power of the player B and decreases negative negotiations power of the player A.

Conclusions from above relations are that by recommending strategy h^* regulator can:

- Increase positive and at the same time decrease negative negotiations power of both players. It occurs when: $V^A(h^*) > V^A(h_{pn})$ and $V^B(h^*) > V^B(h_{np})$.
- Decrease positive and at the same time increase negative negotiations power of both players. It occurs when: $V^A(h^*) < V^A(h_{pn})$ and $V^B(h^*) < V^B(h_{np})$.
- Increase positive and negative negotiations power of player A and at the same time decrease positive and negative negotiations power of player B. It occurs when: $V^A(h^*) > V^A(h_{pn})$ and $V^B(h^*) < V^B(h_{np})$.
- Decrease positive and negative negotiations power of player A and at the same time increase positive and negative negotiations power of player B. It occurs when: $V^A(h^*) < V^A(h_{pn})$ and $V^B(h^*) > V^B(h_{np})$.

As we see, depending on the values of the payoff functions of the players for a recommended strategy h^*, the negotiations power of the players can be changed in different ways. In special cases the regulator can increase or decrease simultaneously positive and negative negotiations power of a one player, or increase one and decrease the other.

In the telecommunications services market, recommended by a regulator strategy h^* represents necessity for interconnection networks of the players on the basis defined by h^*. As we see now, depending on the relations between $V^A(h^*)$ and $V^A(h_{pn})$, and between $V^B(h^*)$ and $V^B(h_{np})$ this necessity may be profitable for one player or both of them or unprofitable for the other or both of players. If h^* were worse[1] for a player than its BATNA, then it would be better for him not to interconnect its network. It is strongly possible that such a situation takes place in the case of incumbent operator. In many situations incumbent is not willing to interconnect its network with the network of a new entrant, because its BATNA (strategy outside the negotiations table) is better than any strategy which could be accepted by the entrant during the negotiations (at the table). At the other side we can expect that usually h^* is better than BATNA of a new entrant, because such operator is willing to interconnect its network on the basis defined by recommended strategy. So we can suppose that in most real

[1] A separate problem arises with an issue of the evaluation's period: short or long?

situations we meet with the last two, above mentioned situations: introducing strategy h^* increase positive and negative negotiations power of one player (new entrant) and simultaneously decrease positive and negative negotiations power of the other (incumbent operator).

Intuitive thought is that it could be the best situation if introducing h^* could improve BATNA of both of the players: $V^A(h^*) > V^A(h_{pn})$ and $V^B(h^*) > V^B(h_{np})$. However we can be sure that it would be the best only if the players would like to agree on the basis of the strategy h^*, because the better h^* the more then they obtain. It in fact means that the players did not agree at the negotiations table, but one of them require arbitration from the regulator. Such arbitration (strategy h^*) would be better for both of players then their BATNA (their best alternative without interconnection) but it have not to be better than the best, accessible, but probably difficult to find solution at the negotiation table. So in some cases we can expect that the better h^*, the more simple negotiations (the players can simply agree to choose h^* at the negotiations table), but at the same time (probably), the more difficult (the less incentive) to find an efficient solution.

From the other side, if h^* were worse than BATNA of both of players: $V^A(h^*) < V^A(h_{pn})$ and $V^B(h^*) < V^B(h_{np})$ interconnection on the basis of the strategy h^* makes a loss for both of players. In other words it would be better for both of them to not interconnect their networks, than interconnect on the basis of h^*. However it does not men that recommendation of h^* worse than BATNA of both of players has not any sens. We should remember that new entrant would like to agree with incumbent on the basis of any strategy that is not worse than its BATNA, and that probably such strategy exists. If such strategy would be also better for incumbent than strategy h^* can be used by an entrant as a threat: "if you don not agree on interconnection on the basis better than my BATNA I would require arbitrations form a regulator!"

What's more, it is also possible, that such situation can give strong incentive for both of players to search for strategy that would be better then BATNA of both of them, so more, probably efficient. So we can expect that recommendation h^* that is worse then BATNA of both of players can make negotiations more difficult, then in the case when h^* were better then BATNAs, but such situation can give more incentive for searching for an efficient solution. However this theoretical conclusions requires verifications from realistic case studies.

Obviously it is also possible, that for the case when h^* is worse than BATNA of both[2] of players ($V^A(h^*) < V^A(h_{pn})$ and $V^B(h^*) < V^B(h_{np})$) strategy h^* can be used as a really antagonistic strategy. One of the players, that aims at deteriorating the payoff of the other player can require an arbitration, because strategy h^* could the most deteriorate the outcome of the other player. Probably it is the most important reason, why it is better from a regulator point of view to recommend h^* that is better than BATNA of both players, or at least which is better than BATNA of a new entrant.

[2]It is important to notice, that if h^* were worse than BATNA of only one of the players, then choosing this strategy generally should not be (though could be) an antagonistic move, because a player, that required an intervention from a regulator may really want to obtain the highest payoff for himself, and that was why he wanted an intervention. But in fact it could be an antagonistic strategy h^* would not be efficient (if both of players could obtain more by choosing different strategy).

4. The impact of recommended solutions on the retail decisions of the players

Existing of a regulators recommended strategy h^* as a reference solution of the negotiations on the wholesale market influences not only the negotiations process (by integration of BATNA, and changing the negotiations power of the players), but also simplifies the process of making a decision on the retail market, that precedes[3] the negotiations. Retail decisions are a part of a whole game, consisted also of wholesale decisions. Recommended strategy h^* simplifies and makes more predictable the process of negotiations. Strategy h^* defines new BATNA of the players and so defines also an integrated reference point, that can be used for pointing out the fair and efficient solution of the negotiations process. This solution can be with higher than without recommendation probability predicted. And so a retail decision that precedes a negotiations is simpler and the finale result of a game more predictable.

With using the concepts of negotiations we can say that existing of a regulator's recommended strategy h^* simplifies the process of structuring the negotiations process or formulating of the problem [16]. By the structure of negotiations we mean here the size of the „cake", and the fair principle of dividing it. The size of a cake can be defined in many ways. One of the possible definitions formulates it as an average value of all accessible and better than BATNA values of payoff function. For example from an A point of view to calculate the size of a cake we should sum up the values of payoffs for such strategies h_l (the strategies accessible in negotiations) that gives the player A payoffs $V^A(h_l)$ higher than his BATNA (than $V^A(h_{pn})$ but also that gives the player B payoffs $V^B(h_l)$ better than his BATNA (than $V^B(h_{np})$). So the size of a cake is calculated with using the following relation:

$$\sum_{l, V_i^A(h_l) > V_i^A(h_{pn}), V_i^B(h_l) > V_i^B(h_{np})} V_i^A(h_l).$$

However, as it was said earlier, it is possible that the player A does not know the BATNA of the player B (the value of $V_i^B(h_{np})$), and so he can not say for which strategy h_l the relation $V_i^B(h_l) > V_i^B(h_{np})$ is true. So during

[3]There is no any impact of the recommending strategy h^* on the retail decisions that are made after a negotiations process, because when such decisions are made, important is only a finale solution of a negotiations and not the way (with or without a regulation) in which this solution was obtained.

calculation the average value of his outcomes, he would have to sum up values for which only relation $V_i^A(h_l) > V_i^A(h_{pn})$ is fulfil:

$$\sum_{l, V_i^A(h_l) > V_i^A(h_{pn})} V_i^A(h_l).$$

So, during calculation the size of a cake to much outcomes $V_i^A(h_l)$ will be used, and it decrease accuracy of this calculation.

Recommendation of the strategy h^* changes this situation. BATNA of both players is defined by this strategy and their are commonly known. So during calculation the size of a cake (an average value of the outcomes) player A can use the relation:

$$\sum_{l, V_i^A(h_l) > V_i^A(h^*), V_i^B(h_l) > V_i^B(h^*)} V_i^A(h_l),$$

what gives higher accuracy of this calculation.

Similarly, existing of the recommended strategy h^* simplifies prediction of the possible, fair principle of dividing the cake. This principle can be defined as a proportion of BATNAs of the players. Without h^* BATNA of the player B could be unknown to player A, and so unknown was the principle of division. Existing of the recommended strategy h^* leads to the situation, that player A can assume that the proportion of the finale outcomes would be mostly near to the $\frac{V^A(h^*)}{V^B(h^*)}$.

Finally a decision problem of choosing retail strategy before negotiations can be formulated as the following optimization problem:

$$\hat{a} = \arg\max_i \left\{ \frac{V_i^A(h^*)}{V_i^B(h^*)} \cdot \sum_{l, V_i^A(h_l) > V_i^A(h^*), V_i^B(h_l) > V_i^B(h^*)} V_i^A(h_l) \right\}, \tag{18}$$

where i is the index of retail strategies a_i.

5. Decision problems of a regulator

The aim of the regulator is to promote competition, and efficiency of the whole market. So the regulator should not provoke, or even create attractive conditions for an antagonistic playing. This is the main reason why h^* worse than BATNA of one or both players should not be referenced. However there are three important problems:

1. It may be difficult for a regulator to get information on the payoff functions of the players, and so difficult to determine the values of outcomes for different strategies.

2. It may be difficult for a regulator to obtain information on a real BATNA of the players.

3. Only wholesale market is regulated, and because of independent decisions on the retail markets the final result of a game even for choosing recommended strategy h^* can be difficult to predict.

First problem, is really a problem if a regulator would like to support of the players in realizing their own aims. If he does not know the payoff functions of the players he could not efficiently support them in realizing the aims that are described by this functions. However this problem is smaller, if a regulator ignores this aims, and is interested only in realizing his own aim, like a desirable market share of both players. For realizing such aim it is not necessary to know the payoff function of the players. However it could be necessary if the players did not want to cooperate on the basis of h^* but only would like to treat this strategy as a reference point (in the sens of proportion of BATNA) in searching different solution during the negotiations. In such a case, finally chosen strategy could not be good from a regulator's point of view. This problem could be partially resolved by waiting with recommending strategy h^* until one of the players requires an arbitration, and than by join in the mediation process, during which regulator could get some important information in an interested matter. Whoever such a situation can never occur. If so, from the regulator's point of view it would be better to give a reference of h^* before a starting of negotiations with hope, that finale result will be close to it.

Obviously, if the players were sure, that a regulator would be interesting in realizing their aims, then it would be profitable for them to inform a regulator about their payoff functions and the aims, they wish to realize.

In the case of unknown BATNA of the players, it is possible that recommended strategy h^* can be worse than such BATNA, and so can be treated as an antagonistic strategy. This problem arises not only from unknown best alternative of the players, but also from unknown payoff functions of the players. Alternatives are evaluated by the values of payoff functions. This payoff functions give an answer on the question why such alternative is the best. So resolving of the problem of unknown BATNA requires first resolving of the problem of unknown payoff functions. However there is also one more problem with unknown BATNA: the higher BATNA, the higher positive negotiations power. So the players want to have as strong BATNA as possible. But the real problem results not from the fact that they want to *have* strong BATNA, but from the fact that they want to make, that the other player think that they have it high even if they really had not. So it is very likely that they would misrepresent in this matter – misrepresent not only in relation to the other player but also in relation to the regulator. So we can expect that in most realistic situations regulators would not know the true BATNA of the players.

The problem of unregulated retail markets arises when a decisions on these markets are made after a negotiations on the (regulated) wholesale market. In such situations recommended strategy h^* determines not a finale result of the whole game, but a vector of possible results (in the case when only one player would make a retail decision after the negotiations) or a matrix of such results (in the case when both players would make retail decisions after the negotiations). In such cases (with assumption that the regulator knows the payoff functions of the players) a decision problem of a regulator should be treated as a multi-objective. Especially the regulator should aim at:

- Maximizing of $V^{A\max}(h^*)$ and $V^{B\max}(h^*)$ for the case of *individually effective* playing on the retail markets.
- Maximizing of $V^{A\min}(h^*)$ and $V^{B\min}(h^*)$ for the case of *antagonistic* playing on the retail markets.
- Minimizing of the coefficients of incentive for playing in an antagonistic way: $\Upsilon^A(h^*)$ and $\Upsilon^B(h^*)$,

$$\Upsilon^A(h^*) = \frac{V^{B\max}(h^*) - V^{B\min}(h^*)}{V^{A\max}(h^*) - V^{A\min}(h^*)}, \qquad (19)$$

$$\Upsilon^B(h^*) = \frac{V^{A\max}(h^*) - V^{A\min}(h^*)}{V^{B\max}(h^*) - V^{B\min}(h^*)}, \qquad (20)$$

where $V^{A\max}(h^*)$, $V^{B\max}(h^*)$, $V^{A\min}(h^*)$ and $V^{B\min}(h^*)$ are the highest and the lowest value of payoffs of the players in the proper vector or matrix.
So choosing of the strategy h^* can be expressed as the following multi-criteria optimization problem:

$$h^* = \arg\max_l \Big\{ [V^{A_{\max}}(h_l), V^{A_{\min}}(h_l), V^{B_{\max}}(h_l), V^{B_{\min}}(h_l), -\Upsilon^A(h_l), -\Upsilon^B(h_l)] \Big\}. \qquad (21)$$

6. Summary

In many real situations, like on the telecommunications services market, free cooperation of the players with highly different negotiations power is impossible. However this cooperation is often necessary for promotion fair and effective competition. That is why it is important to support weaker player in the negotiations process, by the active action of the third side, like, e.g., regulator of the market. Regulators posses highly effective tool for supporting negotiations process: possibility of forcing a recommended solution, strategy h^* which defines new BATNAs of the players, integrates it and makes it commonly known. By introducing a recommended strategy a regulator can effectively change the positive and negative negotiations power of the players. By changing the strategy h^* regulator can:

- Increase positive and at the same time decrease negative negotiations power of both players. It occurs when: $V^A(h^*) > V^A(h_{pn})$ and $V^B(h^*) > V^B(h_{np})$.
- Decrease positive and at the same time increase negative negotiations power of both players. It occurs when: $V^A(h^*) < V^A(h_{pn})$ and $V^B(h^*) < V^B(h_{np})$.
- Increase positive and negative negotiations power of player A and at the same time decrease positive and negative negotiations power of player B. It occurs when: $V^A(h^*) > V^A(h_{pn})$ and $V^B(h^*) < V^B(h_{np})$.
- Decrease positive and negative negotiations power of player A and at the same time increase positive and negative negotiations power of player B. It occurs when: $V^A(h^*) < V^A(h_{pn})$ and $V^B(h^*) > V^B(h_{np})$.

However there are three important problems:

1. It may be difficult for a regulator to get information on the payoff functions of the players, and so difficult to determine the values of outcomes for different strategies.
2. It may be difficult for a regulator to obtain information on a real BATNA of the players.
3. Only wholesale market is regulated, and because of independent decisions on the retail markets the final result of a game even for choosing recommended strategy h^* can be difficult to predict.

This is why the final result of the regulation can be difficult to predict for the regulator. For making a better decision a regulator should get as more information as possible, and precede the decision by multi-criteria analysis of the problem. It is important that in some cases, when the players were sure, that a regulator would be interesting in realizing their aims, it would be profitable for the players to inform a regulator about their payoff functions, alternatives and the aims, they wish to realize. So in such cases it is the challenge for a regulator to convince the players that it would be profitable for them to pass a relevant information.
In the case of telecommunications market, existing of a recommended strategy h^* (independently on its value) ensures that players interconnect their networks. This statement is confirmed by the observation of a market. From theoretical point of view this can be true even if defined by strategy h^* new BATNA's would be weak (in this case the players would have strong incentive to find effective solution). However for reason of possibility of using it as an antagonistic strategy it would be better to recommend such strategy, that defines possibly high BATNA of the players.

Existing of a reference solution on the wholesale market simplifies also preceding the negotiations a decision problem of choosing retail strategy, by increasing the accuracy of the formulating the structure (the size of a cake and the fair principle of dividing it) of the following negotiations.

References

[1] M. Carter and J. Wright, "Bargaining over interconnection: the clear-telecom dispute", Working paper 13, Centre for Research in Network Economics and Communications, University of Auckland, 2000.

[2] N. Economides, G. Lopomo, and G. Woroch, "Strategic commitments and the principle of reciprocity in interconnection pricing", Discussion paper EC-96-13, Stern School of Business, New York University, 1996.

[3] N. Economides, G. Lopomo, and G. Woroch, "Regulatory pricing rules to neutralize network dominance", Industrial and Corporate Change, 1996.

[4] R. Fisher, W. Ury, and B. Patton, *Dochodząc do TAK – negocjowanie bez poddawania się*. Warszawa: Polskie Wydawnictwo Ekonomiczne, 2000 (transl. – in Polish).

[5] K. Fujiwara, “A Stackelberg game model of dynamic duopolistic competition with sticky price”, *Econom. Bull.*, vol. 12, no. 12, pp. 1–9, 2006, http://economicsbulletin.vanderbilt.edu/2006/volume12/EB-06L10030A.pdf.

[6] A. Groebel, “Accompanying the evolving competition: regulatory interconnection decisions and their impact on the telecoms market”, in *Proc. Third Conf. Int. Polish Exper. Interconn. Iss.*, Warsaw, Poland, 1999.

[7] B. E. Hermalin and M. L. Katz, “Network interconnection with two-sided user benefits”, Working paper, University of California, 2001.

[8] G. Iyer and J. M. Villas-Boas, “A bargaining theory of distribution channels”, *J. Market. Res.*, vol. XL, pp. 80–100, 2003, http://groups.haas.berkeley.edu/marketing/PAPERS/IYER/bargaining.pdf

[9] S. Laskowski, “Wspomaganie procesu ustalania cen detalicznych i negocjacji stawek rozliczeniowych na konkurencyjnym rynku usług telekomunikacyjnych”, Warszawa: Politechnika Warszawska, Wydział Elektroniki i Technik Informacyjnych, 2004 (Ph.D. thesis – in Polish).

[10] X. Mancero and E. Saavedra, “Entry, cream skimming, and competition: theory and simulation for santiago de chile’s local telephony market”, Techn. Rep., Instituto Latinoamericano de Doctrina y Estudios Sociales, 2001, www.ilades.cl/economia/publicaciones/ser_inv/inv132.pdf.

[11] H. Raiffa, *The Art and Since of Negotiation*. Cambridge: Harvard University Press, 1982.

[12] B. R. Raineri, “Network competition: a general equilibrium analysis”, Techn. Rep., Encuentro de la Sociedad de Economia de Chile, 2003, http://www.econmeetings.cl/pdf/sesion111raineri.pdf

[13] V. Skreta, “Interconnection negotiations between telecommunication networks and universal service objectives”, Techn. Rep., University of Minnesota, 2002, http://www.econ.umn.edu/ screta/data/net.pdf

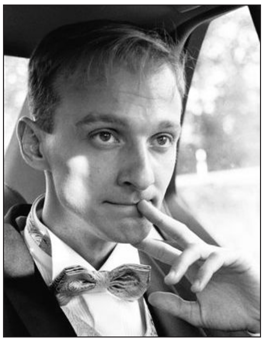
[14] W. Ury, *Odchodząc od NIE – negocjowanie od konfrontacji do kooperacji*. Warszawa: Polskie Wydawnictwo Ekonomiczne, 2000 (transl. – in Polish).

[15] T. Valletti, “Obligations that can be imposed on operators with significant market power under the new regulatory framework for electronic communications”, Techn. Rep., 2003, http://www.ms.ic.ac.uk/tommso/mobile.pdf

[16] M. Watkins, *Sztuka negocjacji w biznesie – innowacyjne podejścia prowadzące do przełomu*. Gliwice: Helion, 2005 (transl. – in Polish).

[17] A. P. Wierzbicki and M. Makowski, “Multi-objective optimization in negotiation support”, Techn. Rep., International Institute for Applied Systems Analysis, Laxenburg, Austria, 1992.

[18] G. A. Woroch, “Local network competition”, in *Handbook of Telecommunications Economics*, chapter 15, Cave and Majumdar and Vogelsang, 2002, http://emlab.berkeley.edu./users/woroch/



Sylwester Laskowski received his M.Sc. (1999) in telecommunications from the Institute of Telecommunications of Warsaw University of Technology, M.A. (2003) in classical guitar from the Warsaw Academy of Music, and Ph.D. (2006) in telecommunications from the Institute of Control and Computation Engineering of Warsaw University of Technology, Poland. His Ph.D. thesis titled “Detail price setting and interconnection fees negotiation process support on competitive telecommunications services market” illustrates application of methodology of a game theory and multi-criteria analysis into strategic analysis and decisions and negotiations support on competitive telecommunications market. His current research interests concentrate on developing this subject matter. He is employed as a Research Assistant at National Institute of Telecommunications.

e-mail: S.Laskowski@itl.waw.pl
National Institute of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland

Paper

Cross-selling models for telecommunication services

Szymon Jaroszewicz

Abstract—Cross-selling is a strategy of selling new products to a customer who has made other purchases earlier. Except for the obvious profit from extra products sold, it also increases the dependence of the customer on the vendor and therefore reduces churn. This is especially important in the area of telecommunications, characterized by high volatility and low customer loyalty. The paper presents two cross-selling approaches: one based on classifiers and another one based on Bayesian networks constructed based on interesting association rules. Effectiveness of the methods is validated on synthetic test data.

Keywords— cross-selling, telecommunication service, classifier, association rule, Bayesian network.

1. Introduction

The definition of cross-selling (according to Wikipedia) is:

"*Cross-selling is the strategy of selling other products to a customer who has already purchased (or signaled their intention to purchase) a product from the vendor*."

Cross-selling offers several advantages. Except for the obvious from the extra products sold, it also increases the dependence of the customer on the vendor and therefore reduces churn. We will now discuss some of the specific aspects of cross-selling in the telecommunication industry, with special focus on cellular network operators.

Telecommunications markets are characterized by high volatility. Customer loyalty is at a very low level in this sector, due to anti-monopoly measures taken by governments, as well as lucrative offers for *new* customers from most service providers.
Cross-selling is thus very important for cellular operator since the more services a user has activated the closer he/she is tied to the company, and the harder it is for him/her to switch to another provider.
In case of telecommunication companies, there exist several marketing communication channels through which a customer can be reached:

– offers made to customer when *he/she* contacts the call-center;

– a phone call to the customer;

– an SMS sent to the customer;

– a standard mail sent to customer (may accompany the monthly bill).

It may seem that some of those channels (especially SMS messages) incur almost no cost, so a large number of offers should be sent. In reality this is not true. The reason for that is the negative reaction of customers to too many offers [3]. Too many SMSes are simply annoying, the users quickly learn to ignore them.
It follows that the amount of cross-selling opportunities is in fact quite limited, and the campaigns have to be carefully targeted such that the probability of a "hit" is maximized.

Let us now briefly discuss related literature and available commercial cross-selling solutions.
A cross-selling application applied in the banking sector is presented in [3]. The system selects customers who would potentially be interested in opening a brokerage account. A classifier (decision tree) is built separately for each service. If a customer who does not have a brokerage account falls into a leaf of the tree where many customers have such an account, it is assumed that the customer is likely to accept the offer. The authors claim that the acceptance rate was much higher than for random offers.
In [15] association rules and statistical models are used to predict purchases based on WWW logs. Association rules are used to generate features which are then used as inputs to a hybrid classifier model.
In [10] the authors present a probabilistic model with hidden variables for predicting customer behavior based on their purchases and questionnaire data. An advantage of such models is high flexibility and possibility of inclusion of hidden variables. A disadvantage is the difficulty of detecting relationships not included in the model. In this work this problem has been solved through the use of association patterns to discover new relationships.
Wong and Fu [18] present a method of selecting a subset of services which should be promoted in order to maximize overall profit. The influence of popularity of some services on the popularity of others is taken into account. The analysis of dependencies between products is achieved through market basket analysis (association rules). It has been shown that selecting an optimal set of products in NP-complete, thus an approximate algorithm has been presented.
A number of companies offer cross-selling products, some of them targeted specifically at telecommunication market. We will briefly describe two such products.
Single attachment station (SAS) offers a telecommunication cross-selling solution [14]. Detailed information is not available, however, the company does say that it is based on market basket analysis [2]. Association rules are used to analyze typical paths of customers' development, e.g., be-

ginning with a single phone line and later moving to a few phone lines plus an Internet connection. This allows for identification of customers who can be interested in purchasing new services. The system is custom built by SAS specialists and requires the purchase of SAS licence.

IBM offers `IBM Guided Selling & Active Advisor` a complete cross-selling solution targeted primarily towards retail sales. No description is available of methods and algorithms used.

Related to cross-selling are so called *recommender systems* [1], which offer suggestions to customers based on the similarity of their purchase histories to histories of other customers. Probably the best known example is the webpage of the `www.amazon.com` online bookstore displaying an information "customers who bought this book also bought...". Such systems are dedicated to retail stores with thousands of products. Telecommunication markets are quite different in this respect since the number of services is much smaller. Also, more data about customers such as sex, calling history, etc., are available, which is not the case for recommender systems. Such systems are thus not very useful for cross-selling in the telecommunication industry.

This paper presents an analysis of two approaches to cross-selling in a telecommunications setting.

The first approach is based on constructing a Bayesian network representing customer's behavior and using this network to predict which customers are most likely to pick each service offered. This gives not only a cross-selling model, but also allows the analyst to gain insight into the behavior of customers. The Bayesian network is constructed using a method based on author's previous work. The method starts with a (possibly empty) network representing users background knowledge. At each iteration patterns are found whose probabilities in customer data diverge most from what the network predicts. The analyst then explains those discrepancies by updating the network.

The second approach uses a separate classifier model for each service offered. Each model predicts, which customers are most likely to buy a specific product. Each customer is then offered a service the classifier of which gives the highest probability of acceptance (among the services which the customer does not yet use). The method does not give any insight into customer behavior but is fully automatic.

2. Test data and experimental setting

Unfortunately the author was not able to perform the experiments on real customer data. Instead, a synthetic data generator developed at the National Institute of Telecommunications was used. Efforts have been made to ensure that the simulation is realistic. To ensure objectivity, the data generator was created by a different person then the one doing the experiments. The experiments revealed that the method based on Bayesian networks achieved lower cross-selling accuracy than the classifier based method, but offered valuable insight into customer behavior (e.g., it was able to reconstruct much of the data generator's internal logic). Below we describe the experimental setting and the data generator.

It has been assumed that the cross-selling action targets three optional services allowing the customer to lower connection cost. The services are described below:

- **RL: cheap local calls** lower price for calls made within customer's local area;
- **TPG: cheap late calls** lower price for calls made after 6 p.m.;
- **TPWS: cheap call within the network** lower price for calls to other users of our network.

The goal is to design a system which for a given customer will suggest one of the above services, which the customer is likely to accept.

Data generator. The generator works in three stages. First customer billing data are generated. Based on those data and a set of rules, active services are chosen for each customer. Data is then aggregated to obtain the format used in data warehouses, e.g., one record of the aggregated dataset corresponds to one customer.

Table 1
Characteristics of customer profiles

Profiles	Characteristics
1	More SMS-type services 60–70% of all uses, short connection time – most connections take just few minutes
2	Most connections during peak hours, few connections outside peak hours
3	Many peak hours connections both within the network and to land lines, less evening connections
4	Most calls during peak hours to land lines, 1 or 2 area codes
5	Most calls to just a few selected users, most calls withing the network

To represent customer diversity, the simulated customers have been split into several profiles. Each profile has different calling habits. Table 1 shows brief characteristics of the profiles.

For each customer we also select at random (taking into account customer's profile) one of six calling plans. In general, the longer a customer talks, the higher plan he is assigned (meaning higher monthly payment but lower cost per call).

Based on customer's profile his/her billing data are generated. Data is then aggregated into a data warehouse format, and services used by each customer chosen based on probabilistic rules. Attributes of aggregated data are given in Table 2.

Table 2
Attributes of the aggregated database table (data warehouse)

Attribute	Description
user-id	User identifier
czas-polaczen	Total connection time
ilosc-polaczen	Number of connections made
sr-dlug-pol	Average connection length
pd_*y*-il-pol	Number of connections during part of day *y*
pd_*y*-czas-pol	Total connection time during part of day *y*
usl_*y*-czas-pol	Total connection time for service *y*
usl_*y*-il-pol	Number of connections for service *y*
taryfa	Customer's calling plan
rl	Active "cheap local calls"
tpg	Active "cheap late calls"
tpws	Active "cheap call within the network"

The following types of calls are available:

- SMS,
- connection within the network,
- connection to another cellular network,
- connection to a land line.

Each day is split into the following three "parts of day": 8:00 – 17:59, 18:00 – 23:59, 0:00 – 7:59.
Numerical variables have been discretized using the equal weight method. About five thousand data records have been generated. The data has been split into the training (4000 records) and testing (1000 records) sets. Models are built on the training set, and their accuracy is verified on the test set. This minimizes the risk of overfitting where the model "learns" the training data but cannot generalize to new examples.

3. Association rules based approach to cross-selling

In this section we describe the association rules based approach to cross-selling. Association rules have first been introduced by Rakesh Agrawal and his team [2] and used to analyze supermarket purchase data. Thus the approach is also known as *market basket analysis*.
Initially association rules have been defined for binary tables, where each attribute corresponded to an item and each record to a transaction. The attribute was set to 1 in a given record if the corresponding item was purchased in the corresponding transaction.

Let $H = \{A_1, A_2, \ldots, A_n\}$ be the set of attributes. Take any subset $I = \{A_{i_1}, A_{i_2}, \ldots, A_{i_k}\} \subseteq H$. The *support* of the set of attributes I in a database table D is defined as

$$\text{support}_D(I) = \frac{|\{t \in D : t[I] = (1, 1, \ldots, 1)\}|}{|D|}, \qquad (1)$$

that is, as the fraction of records in which all attributes in I are simultaneously 1.
If $I, J \subset H$ and $I \cap J = \emptyset$, we can define an *association rule* $I \rightarrow J$. For such a rule we define two quantities which assess its quality: *support* and *confidence*, given by the following formulas:

$$\text{support}_D(I \rightarrow J) = \text{support}_D(I \cup J), \qquad (2)$$

$$\text{confidence}_D(I \rightarrow J) = \frac{\text{support}_D(I \cup J)}{\text{support}_D(I)}. \qquad (3)$$

Support tells us what proportion of transactions in the database contain all items in $I \cup J$, and confidence tells us how likely it is that a transaction containing all items in I also contains all items in J.
In [2] the Apriori algorithm has been presented, which discovers all rules with given minimum support and confidence. Minimum support ensures that discovered rules pertain frequently occurring situations, and minimum confidence ensures high predictive value.
Association rules can easily be generalized to multivalued and numerical (through discretization) attributes.
An advantage of association rules is that existing algorithms allow for finding all rules with given parameters, allowing for discovery of high level correlations. A drawback is that usually too many rules are discovered which creates a secondary analysis problem of finding rules which are interesting to the user. One of such filtering methods (developed by the author) has been applied here to the cross-selling problem.

3.1. Finding interesting association rules

As it has been said above, application of association rules requires methods of selecting interesting rules. One of the methods for achieving this task has been developed by the author of this paper (in cooperation with others) and published in [7, 8].
The method is based on taking into account users knowledge of the analyzed problem. The knowledge is represented using a formal model (Bayesian network). Association rules discovered in data which do not agree with what users knowledge predicts are considered interesting. Such rules are than used by the user to update the model, and the algorithm is applied again to find new interesting rules.
User's knowledge is represented using Bayesian networks [6, 9, 12]. Bayesian networks are directed acyclic graphs depicting direct causal relationships between attributes. Vertices correspond to attributes, and edges to direct causal links. Additionally every vertex is labelled with a conditional probability distribution. A Bayesian network completely determines a joint probability distribution over the attributes it described, allowing for inferences based

on that distribution. Figure 1 shows and example Bayesian network.

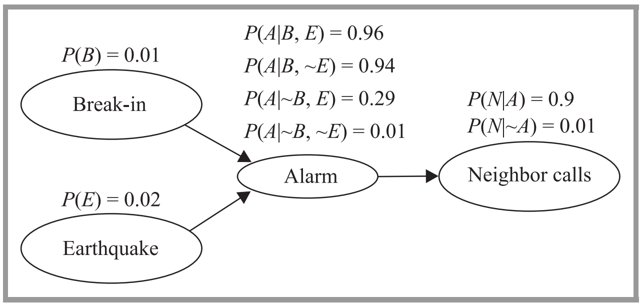


Fig. 1. An example Bayesian network describing simple probabilistic relationships.

One of the main advantages of Bayesian networks is their intelligibility. The dependencies between attributes are shown simply as edges in a graph. Bayesian networks are easy to build, it suffices to connect vertices with appropriate edges. This is usually easy, since humans can easily identify causal relationships [12]. Once the graph has been constructed, conditional probabilities are simply estimated from data. An additional advantage of Bayesian networks is that they determine joint distribution over their attributes, so the description they provide is complete.

Let E be a probabilistic event. The *interestingness* of this event is defined as [8]

$$inter(E) = |P^{BN}(E) - P^{D}(E)|, \quad (4)$$

that is, as the absolute difference between the probability of that event obtained from data and predicted based on the Bayesian network.

Events analyzed in [8] have the form

$$attribute_1 = value_1 \wedge attribute_2 = value_2 \wedge \ldots \wedge attribute_k = value_k, \quad (5)$$

corresponding to sets of attributes in market basket analysis.

The algorithm in [8] finds all such events with given minimum level of interestingness.

One of the main problems related to Bayesian networks is high computational complexity of computing marginal probabilities needed in Eq. (4). Bayesian network inference is NP-complete, and during the course of the algorithm such inference is repeated thousands of times. In Eq. (4) the problem has been addressed by computing larger marginal distributions from the network, and marginalizing several smaller distributions directly from larger ones. This allowed for use of networks of up to 60 attributes. In [7] an approximate, probabilistic algorithm has been given, which works even for huge Bayesian networks, and provides guarantees on the accuracy of discovered patterns.

Detailed description of those algorithms is beyond the scope of this work and can be found in [7, 8].

An important advantage of the approach is that its result is a full probabilistic model, not just a set of rules. The model can then be used for probabilistic inference. Bayesian networks are so flexible, that practically any parameter of the model can be computed from them. This has been used below to estimate the probability of acceptance of a given product by a customer during a cross-selling action.

Adaptations needed for the cross-selling problem. The algorithms described above required certain modifications to work for the given application. Problems occurred when to many edges were directed towards a single node, causing an exponential growth of the conditional probability table associated with the vertex. This caused two types of problems.

The first one was big memory consumption. The second, difficulties in reliable estimation of distribution parameters. The first problem was solved by only keeping nonzero probabilities, the second by using so called Laplace correction to estimate the probabilities. Laplace correction smoothes probability estimates by using a uniform prior distribution.

3.2. Building the Bayesian network

We will now describe the process of building the Bayesian network based on the training set.

Before the first application of the algorithm, edges corresponding to trivial, well known dependencies have been added to the network. These were primarily the consequence of how the attributes were aggregated. Table 3 shows edges in the initial network.

Table 3
Edges corresponding to trivial apriori known dependencies following from the way the data were aggregated

From	To	Justification
pd1-il-pol	ilosc-polaczen	Number of connections is the sum over all parts of day
pd2-il-pol	ilosc-polaczen	
pd3-il-pol	ilosc-polaczen	
usl1-il-pol	ilosc-polaczen	Number of connections is the sum over all services
usl2-il-pol	ilosc-polaczen	
usl3-il-pol	ilosc-polaczen	
usl4-il-pol	ilosc-polaczen	
pd1-czas-pol	czas-polaczen	Total connection time is the sum over all parts of day
pd2-czas-pol	czas-polaczen	
pd3-czas-pol	czas-polaczen	
usl2-czas-pol	czas-polaczen	Total connection time is the sum over all services
usl3-czas-pol	czas-polaczen	
usl4-czas-pol	czas-polaczen	
usl2-il-pol	usl2-czas-pol	Number of connections influences connection time
usl3-il-pol	usl3-czas-pol	
usl4-il-pol	usl4-czas-pol	
pd1-il-pol	pd1-czas-pol	
pd2-il-pol	pd2-czas-pol	
pd3-il-pol	pd3-czas-pol	
ilosc-polaczen	sr-dlug-pol	Average length is computed from total time and number of connections
czas-polaczen	sr-dlug-pol	

Table 4

Results of repeated application of interesting association rule discovery algorithm to the cross-selling problem

Most interesting events					
attributes	values	inter.	P^{BN}	P^{D}	conclusions
First application of the algorithm					
ilosc-polaczen, rl, tpg, tpws	2,N,N,N	0.200	0.1839	0.3845	Number of connections influences additional services used by customers. Customers who make few calls don't use those services. **Added edges** from ilosc-polaczen to rl, tpg, tpws
Second application of the algorithm					
pd2-czas-pol, taryfa, rl, tpg, tpws	1,2,N,N,N	0.179	0.0362	0.2153	Relation between those services seems intuitive. In order to better understand the nature of those relationships, most interesting pairs of attributes were examined
sr-dlug-pol, rl	4,N	0.153	0.2215	0.068	Customers making long calls more ofter use the "cheap local calls" service. The conclusion was considered plausible and **edge has been added** from sr-dlug-pol to rl
sr-dlug-pol, rl	3,N	0.140	0.2077	0.3478	
Third application of the algorithm					
pd2-czas-pol, taryfa, rl, tpg, tpws	1,2,N,N,N	0.18011	0.0351	0.2153	The pattern was still the most interesting one, pairs of attributes were examined again
usl4-il-pol, rl	2,T	0.15	0.1680	0.0183	The influence of the number of calls to land lines on "cheap local calls" is plausible. **Added edge** from usl4-il-pol to rl
taryfa, pd2-czas-pol	1,2	0.1425	0.0727	0.2153	Dependency between calling time during the day and calling plan. **Added edge** pd2-czas-pol to taryfa
Fourth application of the algorithm					
taryfa, rl, tpg, tpws	1,N,N,N	0.151	0.1549	0.3058	Calling plan influences services used. Customers with a cheap plan use their phone infrequently, and thus don't activate extra services. **Added edges** from taryfa to rl, tpg and tpws
Fifth application of the algorithm					
sr-dlug-pol, tpg, tpws	4,N,N	0.149	0.3134	0.1648	Customers making long calls usually have at least one of tpg or tpws active. **Added edges** from sr-dlug-pol to tpg and tpws
Sixth application of the algorithm					
ilosc-polaczen, pd2-il-pol, rl, tpg, tpws	2,2,N,N,N	0.167	0.1068	0.2738	Day time 2 means day time connections so it tells a lot about customer's profile. **Added edges** from pd2-il-pol to rl, tpg and tpws

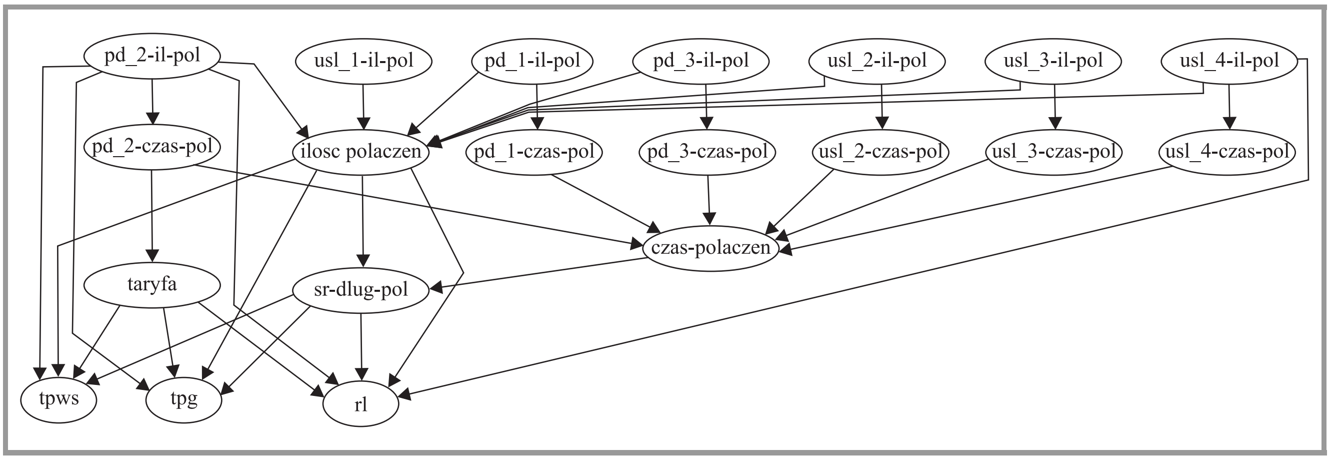


Fig. 2. Final Bayesian network built by analyzing customer behavior data.

Note that the Bayesian network models those dependencies well. For example, connection times in various parts of day are independent of each other. But when the total connection time is known, they become dependent, just as the network predicts.
Table 4 illustrates the process of building the Bayesian network describing customer behavior.
Each subtable shows a new run of the algorithm and the most interesting (in the sense described above) events discovered. The events are conjunctions given in Eq. (5).
The columns of Table 4 are described below:

– attributes: attributes of the interesting event,

– values: values of attributes in the event,

– inter.: interestingness value (Eq. (4)),

– P^{BN}: probability of the event in the Bayesian network,

– P^{D}: probability of the event in the data,

– conclusions: interpretation and explanation of the event, modifications applied to the Bayesian network.

The final Bayesian network is shown in Fig. 2.

3.3. Testing of the model

Reliable testing of a cross-sell solution in off-line conditions is difficult. A real test should involve sending offers based on the analyzed model to a test group of customers, and checking how many of them responded. The procedure should be repeated for another group with random offers. The results for both groups should then be compared.
Such a test was not possible in this work. A simulation of such a test has thus been conducted. We assume that the set of services of a user in the dataset consists of services the user would activate had they been offered (let us call this set A).
For each customer a random subset of those services has been removed (every service was removed with probability 50%). Thus obtained set B was assumed to be the set of services which were active before the marketing campaign. Thus the services in $A \setminus B$ were the services the user would accept if offered.
Then, for each user, based on the Bayesian network the probability of each service not in B was computed, and the offer was made for the service with highest such probability. If the offered service was in $A \setminus B$, the offer was assumed to be accepted. For comparison we also picked a random offer (from those not in B).
The percentage of accepted offers is:

– Bayesian network: 22.84%,

– random offer: 12.83%.

It can be seen that the Bayesian network achieved almost twice as high efficiency as random offers. It should also be noted that in the test set 53.59% of customers did not have any active offer, which in our test prevented them from accepting *any* offer. Since over 50% of offers must have been rejected anyway, 23% accuracy should be considered very high.

4. Classifier based cross-selling approach

In this section we present the second approach which is based on classification models. For each service we want to sell, a classifier is built which assesses the probability that a given customer uses the service. To select which service to offer to a customer we feed his/her data to each of the classifiers and pick the one with highest predicted probability (out of the offers the user does not already have).
As it was mentioned above, this is not an optimal solution. Potential new customers may not resemble current users of

a service. Ideally one should send a pilot offer to a random sample of customers and build classifiers based on the results of that offer. In the current work (and in many real life marketing campaigns) such an approach was not possible.

We have used four classification algorithms implemented in the Weka package [17]: naive Bayesian classifier, decision trees (J4.8), boosted decision trees (AdaBoostM1) and support vector machines. The algorithms have been briefly characterized below, full description is beyond the scope of this work and can be found, e.g., in [17].

Naive Bayesian classifier. Despite being one of the simplest classifier models, this approach often gives results comparable to or even better than other more complicated models [11, 17].

Suppose we want to predict class Y based on attributes $X_1, X_2, \ldots, X_n$. From Bayes theorem we have

$$P(Y = y_i | X_1 = x_{i_1} \wedge \ldots \wedge X_n = x_{i_n}) = \frac{P(X_1 = x_{i_1} \wedge \ldots \wedge X_n = x_{i_n} | Y = y_i) \cdot P(Y = y_i)}{P(X_1 = x_{i_1} \wedge \ldots \wedge X_n = x_{i_n})}. \quad (6)$$

Note that the denominator can be omitted since the probabilities over all y have to add up to one, and we can just rescale the probabilities after classification.

We then use the so called "naive assumption" which says that $X_1, \ldots, X_n$ are independent conditioned on Y, which gives

$$P(Y = y_i | X_1 = x_{i_1} \wedge \ldots \wedge X_n = x_{i_n}) \propto P(X_1 = x_{i_1} | Y = y_i) \cdots P(X_n = x_{i_n} | Y = y_i) \cdot P(Y = y_i). \quad (7)$$

If continuous variables are present, discretization or kernel estimation of conditional distributions is used [11, 17].

Decision trees. Another frequently used classifier model is a *decision tree*. Decision trees [13, 17] are a graphical representation of a decision taking algorithm.

We begin at the root of the tree. In every node we perform an appropriate test and based on its outcome pick the left or right branch of the tree. The procedure is repeated recursively until we reach a leaf of the tree which contains the final decision.

Several decision tree learning algorithms are available in literature [13, 17]. In general the algorithms proceed by picking the test to be placed in the root of the tree, then splitting the dataset in two parts based on the outcome of the test, and then repeating the procedure recursively on each part. After that, the tree is "pruned" to prevent overfitting.

We used the J4.8 algorithm which an improved version of Ross Quinlan's C4.5 method [13].

Boosting. *Boosting* is a method of improving accuracy of other classification models [5, 17]. The idea is based on the fact that classifiers' error can be decomposed into *bias* and *variance* parts. Bias represents classifiers inability to represent complex relationships in data, and variance represents inaccuracies in estimating classifier parameters. In general the lower the bias of an algorithm, the higher its variance.

There exist methods to decrease variance of classifiers. *Bagging* takes several (even hundreds) samples from the training set and builds a classifier on each of them. All those models are then averaged which results in variance reduction.

A better method of variance reduction, which also has the potential to reduce bias is *boosting* [5]. The method works by training a classifier and then reweighting the training set, such that misclassified examples are given higher weights. A new classifier is built on the reweighted data. The process is repeated several times, and all resulting classifiers participate in the final decision. Detailed analysis of the method can be found in [5].

In this work the AdaBoostM1 [5, 17] algorithm was used with J4.8 tree as the base classifier.

Support vector machines. The last classification method is the newest of all four. Support vector machines [4, 16] allow for classification of nonlinear problems while providing guarantees on generalization accuracy for previously unseen cases.

Linear support vector machines construct a linear hyperplane separating both classes. It is constructed in such a way that a large margin between the separating plane and examples from both classes is maintained, which allows for a theoretical guarantee on generalization accuracy.

In order to classify nonlinear problems, original coordinates are transformed in a nonlinear fashion. In the new space the problem may become linear. In order to achieve high efficiency, the transformation is not done explicitly, but achieved through the use of appropriate kernels; see [4] for an excellent introduction.

Experimental results. The classifier based method has been tested the same way as the Bayesian network based model. From among the services the user does not have, we select the one whose classifier gives the highest probability. Effectiveness is estimated exactly as in the previous section. The percentage of accepted offers is:

- naive Bayes: 20.54%,
- decision tree (J4.8): 27.93%,
- AdaBoostM1 (J4.8): 26.19%,
- support vector machine: 28.15%,
- random offer: 12.83%.

It can be seen the naive Bayesian classifier, the simplest classification method, gave results significantly worse than other methods. The remaining three classifiers achieved comparable accuracy, although decision trees have been slightly worse than boosted decision trees and support vector machines.

Almost 30% of offers have been accepted, which means very high effectiveness.

5. Conclusions and further research

It is apparent that classifier based methods achieved (except for the naive Bayesian classifier) higher effectiveness than the Bayesian network.

It should be noted however that such methods do not provide models which are understandable to humans. This is the case even for decision trees, where large size of the tree and potential variable correlations make it difficult to understand the underlying causal structure.

Classifier models can thus be useful for selecting customers and services which should be targeted, but not to explain *why* particular customers prefer particular services. Such knowledge could of course result in a better marketing campaign.

The Bayesian networks based method offers lower accuracy but gives full insight into dependencies between attributes in the data. While building the network we "learn" the data, and eventually get a model describing not just the correlations, but also causal relationships between all variables. We can thus understand how changing one of the parameters will influence probability distributions of other parameters.

The first direction of future research will be improving the Bayesian network implementation such that conditional probability distributions can be represented using classifiers. This should allow the Bayesian network method to achieve accuracy comparable with classifier based methods.

In a longer perspective it would be interesting to create a model which would describe general aspects of customer behavior. It would thus become possible to predict the demand for a service before it was even rolled out to the market. A Bayesian network could form a basis of such model. It would also be useful to couple such a model with customer's lifetime value prediction module.

References

[1] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions", *IEEE Trans. Knowl. Data Eng. (TKDE)*, vol. 17, no. 6, pp. 734–749, 2005.

[2] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases", in *ACM SIGMOD Conf. Manage. Data*, Washington, USA, 1993, pp. 207–216.

[3] M. Berry and G. Linoff, *Mastering Data Mining*. New York: Wiley, 2000.

[4] N. Christianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge: Cambridge University Press, 2003.

[5] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting", Tech. Rep., Dept. of Statistics, Stanford University, 1998.

[6] D. Heckerman, "A tutorial on learning with Bayesian networks", Tech. Rep. MSR-TR-95-06, Microsoft Research, Redmond, 1995.

[7] S. Jaroszewicz and T. Scheffer, "Fast discovery of unexpected patterns in data, relative to a Bayesian network", in *11th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. KDD 2005*, Chicago, USA, 2005, pp. 118–127.

[8] S. Jaroszewicz and D. Simovici, "Interestingness of frequent itemsets using Bayesian networks as background knowledge", in *10th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. KDD 2004*, Seattle, USA, 2004, pp. 178–186.

[9] F. V. Jensen, *Bayesian Networks and Decision Graphs*. New York: Springer, 2001.

[10] W. Kamakura, M. Wedel, F. de Rosa, and J. Mazzon, "Cross-selling through database marketing: a mixed data factor analyzer for data augmentation and prediction", *Int. J. Res. Market.*, vol. 20, pp. 45–65, 2003.

[11] T. Mitchell, *Machine Learning*. New York: McGraw Hill, 1997.

[12] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*. Los Altos: Morgan Kaufmann, 1998.

[13] J. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo: Morgan Kaufmann, 1993.

[14] "Sas cross-sell and up-sell for telecommunications", http://www.sas.com/industry/telco/sell/brochure.pdf

[15] E. Suh, S. Lim, H. Hwang, and S. Kim, "A prediction model for the purchase probability of anonymous customers to support real time marketing: a case study", *Expert Syst. Appl.*, vol. 27, pp. 245–255, 2004.

[16] V. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

[17] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*. San Mateo: Morgan Kaufmann, 2005.

[18] R. C.-W. Wong and A. W.-C. Fu, "ISM: item selection for marketing with cross-selling considerations", in *Eights Pacific-Asia Conf. Knowl. Discov. Data Min. PAKDD*, Sydney, Australia, 2004, pp. 431–440.

Szymon Jaroszewicz received his M.Sc. degree in 1998 from Szczecin University of Technology, Poland, and his Ph.D. degree in computer science in 2003 from the University of Massachusetts at Boston, USA. He is currently an Assistant Professor at the National Institute of Telecommunications in Warsaw, Poland. His research interests include data mining and knowledge discovery, primarily discovering interesting patterns in large databases.

e-mail: s.jaroszewicz@itl.waw.pl

National Institute of Telecommunications

Szachowa st 1

04-894 Warsaw, Poland

Paper

Survey of NGN migration profiles deployed by selected carriers in Europe, North America and Asia

Wojciech Michalski

Abstract— The paper presents the alternative migration profiles to next generation carrier network architecture. It describes the paths of development of fixed public network infrastructure for five selected service providers. The carriers are selected as being in advanced stage of migration towards NGN. Moreover, examples of varying strategies ranging from full PSTN replacement and NGN overlay to constructing an NGN network from the ground up as well as examples of various migration strategies, especially based on IMS and softswitch solutions are presented.

Keywords— NGN, IMS, softswitch based solution, NGN migration strategies, network evolution.

1. Introduction

This article presents profiles of selected carriers as examples of alternative network migration strategies. It describes technical status of carriers's networks when the migration process began, the transformation process and its consequences.

Strategies were selected mainly from the point of view of carrier's goals and influence of factors like cost of maintaining the public switched telephone network (PSTN), competition and development of voice over IP (VoIP) and multimedia services market.

While each operator develops its own unique network migration path, carrier strategies can be categorized into three main groups: full PSTN replacement, new generation network (NGN) overlay and NGN construction step by step from the beginning. Replacement depends on the removal of time division multiplexing (TDM) switches and access infrastructure. It enables seamless transition of plain old telephone services (POTS) users to IP call/session control. Overlay enables migration of subscribers to IP-based, multimedia environment. It includes continued support of existing infrastructure. Carrier's migration paths include some combinations of particular options, especially partial PSTN replacement with NGN overlay. Combination ensures swiftest transformation, but requires significant capital expenditure (capex).

The carriers presented below were selected as examples of above mentioned migration strategies. For purposes of this analysis five carriers were chosen: British Telecom, Slovak Telekom, FastWeb, China Telecom and Verizon, representing three continents: Europe, Asia and North America. By analyzing development of carrier NGN migration strategies it is possible to answer to the question – when will the mass migration to VoIP occur?

2. PSTN replacement – a BT strategy

This strategy is deploying by one of the greatest European carrier – British Telecom (BT). BT is the incumbent telephone service provider in United Kingdom (UK). In June 2004, carrier advertised their converged network migration plan, called the 21st Century Network (21CN). According to the plan, BT declared to build an NGN network over the next six years. The plan assumes replacement of core PSTN network with an IP network based on the IP multimedia system telecoms & Internet converged services & protocol for advanced networks (IMS TISPAN) architecture.

British Telecom began an NGN trial in 2004. First, end-to-end communication was set up between three major network nodes, one located in Cambridge and two in London. In the first phase, 1000 subscribers participated in trials of end-to-end voice and data services over Internet protocol multiprotocol label switching (IP/MPLS) network. In the second step, by June 2005, BT added additional 3000 subscribers to the pilot network.

Now, to support future broadband growth, BT creates an IP/MPLS core network along with the consolidation of central offices and deployment of multiservice access nodes (MSANs). This process will continue during 2009, when nearly all BT subscribers will be served through NGN access nodes. The MSANs nodes will be handling POTS subscribers too, but no special provisioning of equipment at the customer promises or central office is required.

Large scale migration of non-PSTN services to the NGN began in 2007. BT plans that over 50% its current POTS users will have migrated to the 21st Century Network by 2008.

To support NGN broadband services, BT adopted a service creation platform that consists of components such as authentication, directory and profiles, quality of service (QoS) and presence/location. These components can be combined in a modular fashion when new services are created.

2.1. Next step

In the business market, BT implements the migration plan to the next generation services now. But to retain enterprise customers, BT must continue to expend its enterprise NGN infrastructure, so BT will build an overlay network to specifically address the enterprise market.

Operator knows that further profits could be limited by marriage of wireless fidelity (Wi-Fi) and cellular network and by making BT a "transport only" carrier. To defend his position, BT needs to be a major market player providing NGN customer services. To defend his status, it also needs a wireless networks, although there is a large number of cellular competitors in the UK.

2.2. Usefulness of this method

For BT, the total cost of maintaining the PSTN is higher than for other European incumbents, because it possesses an extensive network infrastructure and must fund all maintenance and upgrades on many switches. For this reason, BT chose the strategy of PSTN replacement.

Many big and small incumbents in other countries also prefer the strategy of full PSTN replacement. PSTN replacement has been initiated by many carriers in Asia, Australia and Oceania. Telecom New Zealand has elaborated an aggressive plan to migrate to NGN architecture over the next few years. Similar plan is advertised even in the small country of Brunei. Forward thinking companies have even outsourced network maintenance and management of both its wired and wireless networks.

3. Partial PSTN replacement and overlay – a Slovak Telekom migration strategy

Strategy of partial PSTN replacement and overlay has been chosen by Slovak Telekom (ST), the incumbent telephone service provider in the Slovak Republic.

In the early 1990s, the telecommunication infrastructure in the Slovak Republic was in very bad technical state. But in a few next years, ST has upgraded portions of the network with fiber cable, synchronous digital hierarchy (SDH) transport systems and digital circuit switches. By 2000, ST has digitized over 70% of the PSTN infrastructure.

Modernization process of ST network didn't cover rural networks still served by electromechanical switches. In this situation, delivering services with required QoS parameters was a big problem. ST reported nearly 28 faults per 100 main access lines in 2000.

As a majority owner, Deutsche Telekom initiated actions to fully digitize the ST network by the end of 2004. Under the contract signed in April 2004, Alcatel was obliged to supply next generation infrastructure to replace analog switches and transmission equipment. According to the contract, Alcatel replaced over 300 small analog switches handling over 200 000 subscriber lines with single Alcatel 5020 softswitch ($1+1$ configuration) and over 7500 Media Gateways operating together with other PSTN switches.

At the transmission level, the core network was built using an IP/MPLS technology. The country was divided into three regions: west, central and east and each of them was configured with dual Cisco GS 1200 core routers for network survivability.

Like most incumbent carriers in Eastern Europe, ST has a foreign ownership which plays important role in the migration process. Thanks to this, ST has greater access to investment capital, technical knowledge and operations expertise.

3.1. NGN technologies a chance for countries lacking advanced infrastructure

Development of NGN infrastructure in the Slovak Telekom network is an excellent example of success on the world scale. This example shows us what several incumbent carriers did over a few years. At first glance, the idea that rural areas having low population densities, low computer and Internet penetration as well as low demand for telecommunication services are to be equipped with NGN infrastructure seems illogical. According to popular opinion, primarily a low cost network solutions should be installed in rural areas. Thanks to jump to NGN, Slovak Telekom can minimize network investment, increase number of subscribers in next years and attain important social goals, such as upgrading rural telecommunication services.

Network transformation process will probably last many years, and ST will establish overlay broadband access networks (digital subscriber line (DSL) and wireless) with connectivity to the IP/MPLS core network over this time. ST will deliver VoIP services to drive further PSTN migration.

3.2. Usefulness of this approach

Strategy based on co-existence on NGN and PSTN infrastructure will be common around the world. If a growth of VoIP and other NGN services will be slow enough, the cost of this process shall remain acceptable.

The least developed countries (LDCs) in Africa and Southeast Asia have infrastructures like Slovak Telekom before the upgrades. Therefore, LDCs will likely emulate the ST network migration strategy. They will leapfrog the digital circuit switch technology and replace electromechanical exchanges in rural areas with NGN-IP infrastructure.

4. NGN overlay – a China Telecom strategy

China Telecom is a true monopoly telecommunication carrier serving two thirds of China territory and controlling the national long distance network and provincial networks in 20 provinces, autonomous regions and municipalities.

For European people, it is hard to imagine to growth of China Telecom fixed network over the past decade.

In 1998, China Telecom had 87.4 million total main access lines. By the end of 2003, number of subscribers increased to 263 millions, and by the end 2004 to 299 millions. In 2008, there will be an estimated over 310 million main access lines. China Telecom also has over 40 million wireless users and over 12.5 million DSL broadband subscribers. Although broadband deployment has been growing at the rate of over 200% per year, China Telecom represents only 11% of total residential access lines.

4.1. China Telecom network evolution

The national backbone network of China Telecom is based on fiber-optic cable systems linking provincial capitals. During the 1990s, the transmission network based on fiber cables was upgraded with SDH transport systems, forming rings for improved survivability.

China Telecom PSTN consists of five-level circuit-switched network and all major national transit exchanges are duplicated. The national and provincial PSTNs are relatively new, with age ranging from 10 to 15 years. Age of telecommunication network in China is not a primary factor influencing the necessity of upgrades. The network development is caused by economic growth and increasing subscriber demand. These factors have decisive influence on growth of traffic volume in the network. For this reason, China Telecom started to deploy a next generation overlay network.

Carrier began this processs in October 2004 when it has selected Lucent and Nortel to upgrade its existing SDH national backbone network. Under the contracts, Nortel upgraded existing metropolitan optical networks with wavelength division multiplexing (WDM) in ten major cities and Lucent installed its WaveStar OLS 1.6T high-capacity dense wavelength division multiplexing (DWDM). These contracts show that China Telecom wants to retain the existing TDM-based PSTN network for many years yet.

In November 2004, China Telecom has advertised the award of major contracts for the construction of IP/MPLS core network that are part of the ChinaNet Next Carrying Network (CN2), designed to support IP-based consumer and business services.

In 2002, China Telecom has started softswitch trials with multiple equipment vendors. These trials included Alcatel's, Nortel's, Ericsson's and Lucent's equipment and each of them was dedicated to test system performance, network interworking and equipment inter-operability. Moreover, the goal of the trials was to provide China Telecom with expertise in supporting business VoIP multimedia services.

In July 2003, Shanghai Telecom, a China Telecom's subsidiary, has concluded a contract with Alcatel for construction of next generation metropolitan network. Under this contract, a network was deployed based on Alcatel softswitch, media gateway and litespan multi service access gateway equipment to support integrated IP voice, data and multimedia services over asymmetrical digital subscriber line (ADSL) and Ethernet broadband connections.

4.2. Overlay networks in China Telecom in the past

Overlay networks have been used in China Telecom over the past 25 years. First, digital exchanges and integrated services digital network (ISDN) overlay networks were deployed to upgrade communication services to government and business customers. Than, gradually, the SDH transport systems and fiber-optic cable systems were installed to support the PSTN digitalization.

Overlay networks in China Telecom have been deployed for reasons that remain unchanged to this day. First, the size of China Telecom network causes that deployment of new technologies occurs in phases. Moreover, when this type of deployment is adopted, investment is finished after each phase. Third, under any scenario, replacing about 200 million digital switch lines would take a long time.

Unlike BT, which prefers a next generation network solution that supports the new as well as old services, China Telecom will not adopt old services in its network. Interworking in the existing PSTN network is based on new requirements, but old services, such as narrow-band POTS need not be supported by the next generation network platform.

4.3. Usefulness of this method

Overlay networks permit rapid implementation of new services in selected geographic areas and minimize the risk of disruption to existing PSTN network. This is why China Telecom selected such a strategy, because similarly to operators in other emerging countries, it must meet the immediate demand from businesses for IP-based services.

Establishing overlay networks gives many advantages, but the concept of overlay network has one fundamental disadvantage: no reduction of network cost, because maintaining the existing PSTN infrastructure and building IP/MPLS overlay network is necessary.

5. Construction of NGN step by step from the beginning – a FastWeb strategy

In 1999, FastWeb was established as a joint venture between AEM, Milan's utility and e.Biscom. The company is an Italian broadband telecommunications service provider. Using AEM's underground ducts, FastWeb has built an extensive fiber optic cable network covering Italy's major metropolitan areas.

FastWeb provides dedicated Internet access, voice and video-on-demand (VoD) services, offered to business and residential customers. Thanks to its fiber network, the company was the first European carrier to offer 10 Mbit/s Ethernet-based Internet access to customers (it also uses DSL to provide service beyond its own network coverage).

5.1. Development of FastWeb's network infrastructure

Transformation process was initiated in backbone network, where FastWeb built its fiber-optic transmission network. The core transport network consists of SDH STM-16 dual-fiber, bi-directional ring, supplied by Alcatel. In each core network node, the Cisco 12000 series IP routers are installed. At the customer premises, Cisco's Catalyst switches are used for access and traffic aggregation purposes.

Modernization process of FastWeb network was divided into 5 phases. In phase 1, H.323 as well as access gateways to support VoIP services (translation of E.164 to IP address) were implemented. Call control to each of the access gateway devices as well a supporting endpoint registration is provided through dedicated gatekeepers and the access to the PSTN through the italtel multi service solution (iMSS) softswitch platform.

In phase 2, the softswitch-based service-layer application platform was implemented. This platform has made the provisioning of voice services more efficient and has improved call completion by dynamically rerouting calls upon congestion. In 2002, NetCentrex CCS softswitches were introduced to manage call routing and signaling as well as to offer full set of voice services. FastWeb installs one CCS softswitch cluster for each zone covering about 200,000 users (each CCS cluster provides N+1 redundancy). Moreover, in this phase, a video application platform was added, to support broadcast television, video conferencing and VoD services over its fiber-to-the-home (FTTH) network.

In phase 3, FastWeb deployed session initialization protocol (SIP)-based gateways in its network. Earlier, in phase 2, an Italian access gateway supplier Telsey had deployed its H.323 access gateways. Subsequently, NetCentrex and Telsey worked together to develop a SIP-based access. At the same time, FastWeb installed MSANs, to support the provision of broadcast television services over its ADSL network.

In phase 4, NGN services including server-based applications such as presence and multimedia business services were implemented.

In phase 5, connectivity with other IP networks and application domains were added. FastWeb network evolution composed of 5 phases is not finished yet, because FastWeb still continues to expand network capacity.

5.2. FastWeb network as a VoIP oriented architecture

During FastWeb's network evolution a trial with services expansion was made. In each step of this process consumer and business product as well as services portfolio was added. Thanks to a fiber-optic backbone network, FastWeb has distinguished itself from Telecom Italia and other competitors with high-bandwidth data services. From the beginning, FastWeb has made voice communications a major part of its business model and has recognized that VoIP services will be a fundamental sector of services market to ensure profitability.

5.3. Usefulness of this method

Many other competitive local exchange carriers (CLECs) around the world prefer network migration strategies similar to FastWeb's, because the role and scope of VoIP services will increase in future and NGN architecture will give the carriers the most competitive standing on the voice services market.

6. PSTN replacement – a Verizon strategy

Verizon, being one of the greatest of US carriers, deploys a PSTN replacement without expansion of territorial coverage. Verizon was created in 2000 after a merger of Bell Atlantic and GTE Corporation. Thanks to the merger, Verizon combined the Bell Atlantic operating region with ILEC telephone operations throughout the US midwest, southeast and west. Currently, Verizon operates in 30 states and has also international operations in Puerto Rico and Dominican Republic as well as Vodafone OmniTel in Italy. Moreover, Verizon and Vodafone own Verizon Wireless – the largest wireless carrier in the US.

6.1. Development of Verizon's network infrastructure

Verizon began replacing its long distance national transit (class 3) circuit switches with packet switching technology in 1999. During 2004, the softswitches were installed for class 4 inter-city transit applications. The replacement of primary toll switches and upgrade of the national transit network to IP technology took place in 2003.

In 2003, Verizon established an IP/MPLS backbone network which operates IP over SONET based on Lucent solutions. Under the contract, Lucent supplied SONET equipment for metropolitan network applications and LambaXtreme ultra-long-haul DWDM optical equipment installed in the national backbone network.

In 2004, the company it began deploying fiber-to-the-premise (FTTP) access network in selected states with data speeds of 622 Mbits downstream and 155 Mbits upstream. Moreover, in 2004, under the contract with Nortel, Verizon also began replacing GTD-5 EAX local exchanges.

Verizon deploys FTTP on a regional basis, with complete replacement of copper PSTN access network. Its strategy is a full PSTN replacement rather than establishing a PSTN overlay.

6.2. Usefulness of this method

Public switched telephone network replacement gives Verizon advantages similar like in the BT case. Similar to British Telecom, Verizon found the high PSTN maintenance costs as significant factor in replacing portions of the access network. Moreover, like BT, Verizon will continue support for PSTN circuit switches for many years in the future. For this purpose, Nortel has received contract for the softswitch support and Lucent has received similar contract for 5ESS product support.

Other carriers in Canada and US will probably also adopt similar migration strategies. TELUS covering British Columbia in Canada is expanding its services with a fiber based IP backbone network. Bell Canada is expanding its coverage to provide service across Canada and the US. The combined SBC and AT&T will also deploy its network in similar manner, adopting an in region PSTN replacement and out of region expansion strategy.

7. Conclusion

Examples presented in this article illustrate trends in migration of traditional network to NGN. Now, many carriers in the world build or upgrade their networks to meet requirements imposed by new broadband multimedia and VoIP services.

In Europe, this process was first initiated by BT, which elaborated and introduced network transformation strategy based on IMS system. BT has adopted a full PSTN replacement strategy and had an immediate need to develop an IP/MPLS core network to support both enterprise and consumer services. Currently, BT consolidates its central offices and deploys MSANs to lower operating costs.

Similar strategies, but based on softswitch solution are deployed by other carriers in Europe, e.g., by FastWeb in Italy an by Slovak Telekom in the Slovak Republic. ST has implemented an interesting network transformation scenario promising to go from electromechanical exchanges to NGN infrastructure without the necessity of deploying digital exchanges, employing a hybrid network migration strategy. In future, rather than deploying existing digital circuit switches, it will deploy an NGN architecture with centralized softswitch control and distributed media gateways, based on redundant IP/MPLS core network. FastWeb, being an Italian CLEC, has built an NGN network from the ground up. It was primarily a fiber based bypass service provider, providing broadband services (up to 10 Gbits) to businesses and customers in 14 metropolitan areas. Its network control layer and application servers were introduced via a softswitch deployment.

Carriers in other countries, e.g., France and Germany concentrate mainly on the works concerning planning and elaborating of network transformation strategy suitable to technical characteristics of existing infrastructure and economic standing of the carriers.

In Asia, advanced works concerning transformation of traditional network to NGN infrastructures are carried out by operators in South Korea, China and Taiwan, which, like in Europe, are deploying various scenarios based on IMS and softswitch solution as well as various options of this strategy (PSTN replacement, overlay). China Telecom is an excellent example of a carrier with a network overlay migration strategy. Having PSTN infrastructure about ten years old, China Telecom is constructing an NGN to support the introduction of IP based enterprise services and to expand network capacity for entry into the 3G mobile market.

In North America, Verizon has adopted a migration path based on PSTN replacement in the region and out of the region expansion. It operates with IP as well as circuit-switches services, providing unified messaging, call management and other PC-based applications. It has established a national IP/MPLS backbone network and has introduced web-based, overlay services for managing home and business communications. It replaces GTD-5 exchanges with centralized softswitch-based call control and distributed media gateways and deploys FTTP in selected in-region communities.

Carriers that prefer full PSTN replacement will deliver NGN voice services earlier and more aggressively than those supplementing existing broadband network infrastructure with an NGN overlay. Carriers will develop strategies in accordance with competitive market conditions. Establishing a low-cost, application-rich NGN architecture is a key to a longer term consumer VoIP marketing strategy.

References

[1] "BT perspectives for NGN and standards", http://www.itu.int/ITU-T/worksem/ngn/

[2] "Fastweb FTTx", http://www.point-topic.com

[3] "Slovak Telekom – solution for better life", http://www.eurescom.de

[4] "China Telecom network architecture", http://www.chinatelecom-h.com

[5] "Development scenario of softswitch standards in China and China Telecom's considerations on network evolution", http://www.itu.int/ITU-T/worksem/ngn/

Wojciech Michalski was born in Bogate, in Poland, in 1952. He received the M.Sc. degree in telecommunications engineering from Technical University in Warsaw in 1977. He has been with the Switching Systems Department of National Institute of Telecommunications (NIT) since 1977, currently as a senior specialist. His research interests and work are related to PSTN backbone and access networks, GSM networks and IP networks. He is an author and co-author of technical requirements and many documents concerning telecommunication networks, services and protocols.

e-mail: W.Michalski@itl.waw.pl
National Institute of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland

Paper

Testing of highly doped and photonic crystal optical fibers

Krzysztof Borzycki

Abstract—The paper presents optical measurements – spectral loss, OTDR and PMD, temperature cycling and mechanical tests – bending, twist and crush, performed on Yb-doped single mode fibers and small-core photonic crystal fibers (PCF). Several issues related specifically to characterization of such specialty fibers, like measurement errors and artifacts as well as coupling of test instruments to samples are presented. Of particular importance is reliable and low-loss fusion splicing of specialty fibers to standard single mode fibers (SMF), as most commercially available fiber test instruments are fitted with SMF interfaces only.

Keywords— highly doped fiber, photonic crystal fiber, measurements, testing, polarization mode dispersion, birefringence, mechanical testing, temperature cycling, optical fiber splicing.

1. Introduction

As the development of specialty fibers progresses and their applications in optical amplifiers, fiber lasers, dispersion compensators, sensors, wavelength converters, etc., become more numerous, characterization of such fibers gains importance. Designs, geometry, operating wavelengths and optical parameters of specialty fibers are often very different from those typical for fibers used in telecom networks, for which established measurement techniques, standards and instrumentation have been developed. In addition, new designs are steadily added and standardization is generally lacking. This forces researchers to develop novel testing techniques and ways to test non-standard fibers with existing instruments. An important problem is splicing of specialty fibers to standard single mode or multimode fibers and reduction of splice loss.

All experiments presented in this paper were carried out at the laboratories of National Institute of Telecommunications (NIT) as part of participation in the COST Action 299 "Optical Fibres for New Challenges Facing the Information Society" (FIDES)[1], dedicated to new applications of fiber optics. This includes extensive research and characterization work on new fiber designs, in particular highly doped fibers (HDF) for lasers and amplifiers and photonic crystal fibers (PCF) for sensing and signal processing.

Fiber samples provided by other participants of COST-299 for characterization included:

- Ytterbium-doped, silica-based single mode and multimode fibers intended for high power optical amplifiers, with optical pumping and amplification at wavelengths of 976 nm and 1060 nm, respectively. Those fibers are made commercially by nLight (formerly Liekki Oy), Finland using an unique direct nanoparticle deposition (DND) process and have very high Yb content of 1200 ppm.
- Highly nonlinear "holey", or photonic crystal fibers with small core strongly doped with germanium, supplied by IPHT Jena, Germany.

While the tests were aimed at establishing fiber characteristics such as optical loss, uniformity, polarization mode dispersion (PMD), thermal and mechanical properties, a separate problem of importance was optical coupling between measuring instruments and samples under test.

2. Measurement set-up

2.1. Optical connections to test instruments

Most instruments in our lab, except for optical power meters, had optical interfaces tailored to testing of single- or multimode telecom fibers with standardized cladding diameter of 125 μm and core size of either 5–10 μm or 50–62.5 μm, fitted with optical connectors.

Our preferred approach was to fusion splice the sample of specialty fiber to short (approx. 2 m) lengths of best-matching telecom fiber – either non-dispersion shifted single mode (ITU-T G.652, IEC B1) or 50/125 μm multimode (ITU-T G.651), terminated with FC/PC connectors.

Attempt to connectorize PCF fiber by gluing and polishing has failed: dust created during polishing has filled holes and could not be removed, introducing loss of over 30 dB. Adapters for cleaved fiber could be used, but were available for 125 μm clad fibers only.

2.2. Fusion splicing of PCF to standard single mode fiber

Splicing of single mode highly nonlinear PCF (Table 1 and Fig. 1) was challenging due to several factors:

– mismatch between core size of PCF and standard single mode fiber (SMF);

– different level of GeO_2 doping and refractive index of cores leading to Fresnel reflection;

– holes in PCF being easily be filled with any solvents used to clean the fiber;

– collapse of air holes when PCF is heated.

[1]See, http://www.cost299.org

Table 1
Comparison of PCF and SMF [1, 2]

Parameter	PCF (IPHT Jena 252b5)	SMF (Corning SMF-28)
Cladding diameter [μm]	80	125
Cross-section taken by holes [%]	18.2*	0.0
Cladding diameter after collapse [μm]	72	125
Core diameter [μm]	0.5/2.0**	8.2
Max. refractive index difference [%]	3.85	0.36
Mode field diameter (MFD) @ 1550 nm [μm]	5.8	10.4

* 90 holes of 3.6 μm diameter – see Fig. 1.
** PCF core has a triangular "pedestal" and inner "peak" of step profile.

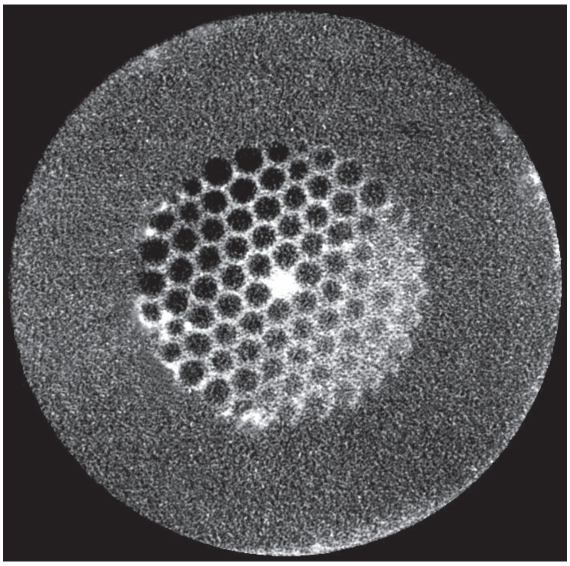


Fig. 1. Cross-section of IPHT Jena 252b5 highly-nonlinear PCF fiber (edited IPHT photo).

First attempts to fuse cleaved PCF and SMF routinely produced a bubble of 20–30 μm size at the interface, as some of the air from collapsing holes became locked there.

Cleaved end of PCF had to be pre-collapsed first (Fig. 2) by low-power electric arc, using current of about 12 mA with electrode spacing of 1 mm; this reduced PCF diameter by 10%. With some 100 μm of PCF tip collapsed, fusion splicing to SMF was made. This included: pre-fusion at 9 mA lasting 4.0 s, fusion at 16 mA with duration of 1.0 s and annealing at 8.5 mA for 3.0 s. To avoid deformation of thin PCF, fibers were put into contact approx. 150 μm away from the axis of electrodes, so most of arc power was transferred to bulkier SMF. Slices were protected with 60 mm long heat-shrinkable sleeves.

While this produced splices of satisfactory appearance and strength, the round trip SMF-PCF-SMF loss was very high: 32–50 dB at 1550 nm. The problem was traced to penetration of PCF holes by solvent (acetone) used to clean the fiber; it decomposed during fusion, leaving a dark, faintly visible carbon residue inside holes. This carbon and solvent locked in the holes produced strong light absorption.

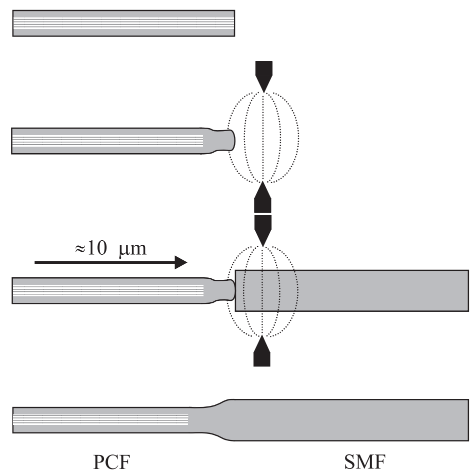


Fig. 2. Fusion splicing of PCF to standard single mode fiber.

In the next attempt, PCF tip was first melted in electric arc and sealed, then coating was softened by acetone bath lasting about 30 s, the fiber mechanically stripped, wiped with acetone-soaked tissue and cleaved. The SMF – 1 m PCF – SMF loss at 1550 nm went down to 15.8 dB. Excluding loss of connector ($\approx$ 0.25 dB) and 1 m of PCF ($\approx$ 0.05 dB), we get 15.5 dB for the SMF-PCF and PCF-SMF splices. Loss spectrum obtained with optical spectrum analyzer (OSA) and tungsten lamp (Fig. 3) shows weak OH^- absorption peak at 1380 nm and some reduction of loss with wavelength, as PCF mode field diameter increases.

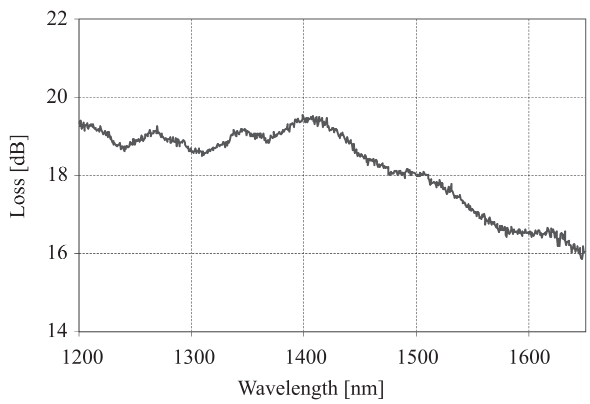


Fig. 3. Spectral loss of 1 m IPHT Jena 252b5 PCF spliced to 2 m long SMF pigtails.

Loss measured in our experiment was far higher than reported in [2], where optimized fusion splicing of identical PCF with collapse and rounding of fiber ends reduced

SMF-PCF-SMF loss to 3.2–4 dB. It was nevertheless acceptable for experiments on short samples.
Use of splicing procedure presented in [2] in our lab, with 0.5 s fusion duration and 18 mA current, followed by 3 extra heatings in the same conditions, gave an SMF – PCF – SMF loss of approx. 2.5 dB at 1550 nm. This work will be presented in a separate paper.
Splicing loss of incompatible fibers can be significantly reduced by introducing a short fiber with intermediate core size and refractive profile [3], but such fiber was not available. Another approach is to keep fusion time very short, typically 0.2–0.4 s [4, 5]. This prevents hole collapse and beam expansion inside PCF, but at the expense of splice strength.

2.3. Fusion splicing of Yb-doped HDF to standard single mode fiber

Two types of double-clad single mode highly doped fibers were tested: a small core (MFD = 6 μm at 1060 nm) Liekki Yb1200-6/125DC [6] and large mode area (MFD = 10 μm at 1060 nm), polarization-maintaining Liekki Yb1200-10/125DC-PM [7]. Both had cladding diameter of 125 μm and no holes, so a standard fusion splicer settings were adopted for HDF-SMF splicing. Fusion time was 1.5 s and arc current 17 mA; it was preceded by 5 s pre-fusion at 9.5 mA and followed by 3 s of annealing at 7.8 mA.
Optical time domain reflectometer (OTDR) traces of SMF to HDF connection show that large part of light entering small core HDF (Yb1200-6/125DC) was forced into inner cladding. Cladding modes disappeared only after 100 m (Fig. 4).
Influence of cladding modes was less pronounced at 1550 nm, which can be explained by smaller MFD difference between fibers at this wavelength. Apparent connection loss resulting from fusion splice and FC/PC connector (on SMF) was 0.3 dB at 1310 nm and fell to –0.1 dB at 1550 nm.
Cladding mode effects and extended dead zone of attenuation measurement were absent in case of large mode area HDF, whose MFD was greater than that of SMF: ≈ 11 μm versus 9.2 μm at 1310 nm (Fig. 5). Apparent connection loss, however, was fairly high: 1.5 dB.

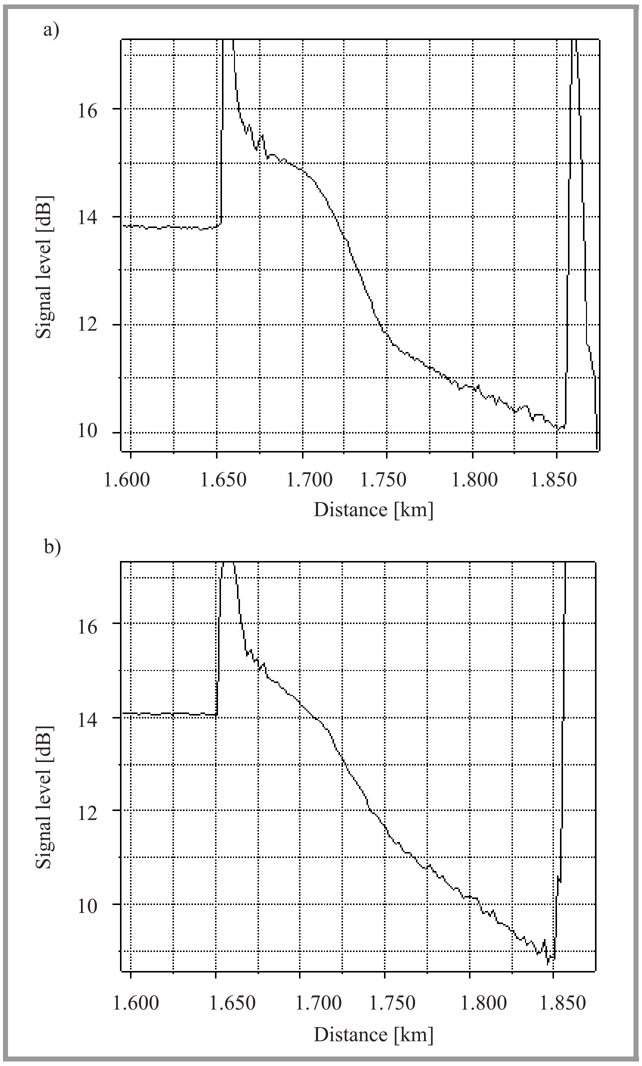


Fig. 4. OTDR traces of Yb1200-6/125DC fiber with SMF pigtail, connected to 1650 m of SMF: (a) $\lambda = 1310$ nm; (b) $\lambda = 1550$ nm.

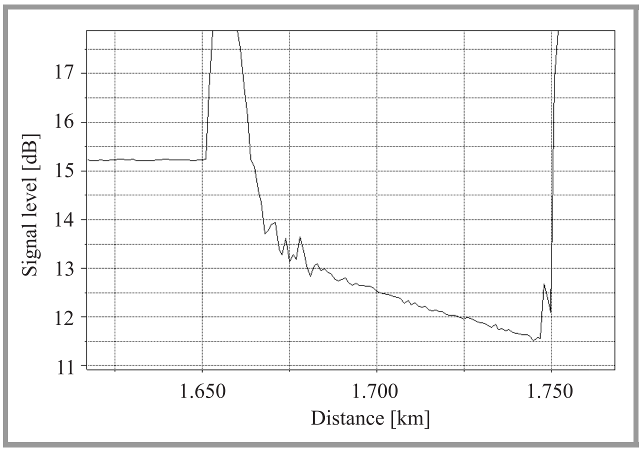


Fig. 5. OTDR trace of Yb1200-10/125DC-PM fiber with SMF pigtail, connected to 1650 m of SMF ($\lambda = 1310$ nm, pulse width 50 ns).

In summary, while fusion splicing of single mode HDF to SMF was easy, possible excitation of slowly-decaying cladding modes was detrimental to optical measurements.

2.4. PMD measurements

Polarization mode dispersion (PMD) in single mode fibers had been measured using alternatively:

– Jones matrix eigenanalysis (JME) method,

– fixed analyzer (FA) method.

Both methods are standardized for telecom single mode fibers [8], but applicability of them and particular instruments to testing of specialty fibers had to be verified. We have also tried to determine dependence of PMD on direction of light propagation.

For JME measurements, we have used an Adaptif Photonics (now Agilent) A2000 PMD analyzer and Agilent HP 8168F tunable laser source. Spectral range and resolution were 1460–1590 nm and 0.001 ps, respectively. The setup also measured other fiber parameters like polarization dependent loss (PDL) or second order PMD and their spectral distribution.

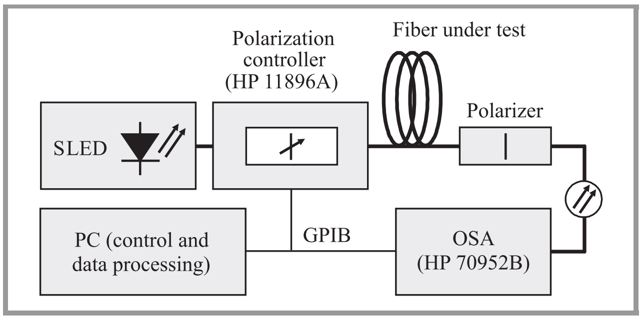


Fig. 6. Block diagram of PMD measurement setup using FA method.

The FA setup (Fig. 6) was assembled using off-the-shelf components. Instruments were controlled by a personal computer (PC) through the general purpose interface bus (GPIB). A superluminescent light emitting diode (SLED) source gave spectral range of 1250–1650 nm and resolution of 0.01 ps. This setup had a short measurement time of 5–20 s. However, FA method based on counting of extrema in transfer characteristics, while reliable, does not provide information on spectral distribution of differential group delay (DGD) or any other polarization-related fiber parameters [8].

3. Test results

Below are presented selected results of measurements and tests performed on HDF and PCF fibers, in particular those highlighting unique characteristics of such fibers and measurement problems encountered when established measurement methods were applied to characterization of those fibers.

3.1. Yb-doped fiber: Liekki Yb1200-6/125DC

Attenuation of this fiber (Table 2) could be measured non-destructively with reasonable accuracy only with OTDR, as this instrument enabled to select fiber section unaffected by propagation of cladding modes (Fig. 4).
Attenuation of this HDF was apparently dictated by Yb_2O_3 doping, not fiber defects. For comparison, loss measured

Table 2
Attenuation of Yb1200-6/125DC fiber – OTDR test

Wavelength [nm]	Attenuation [dB/km]
1310	13.3
1550	26.1

in multimode HDF with the same dopant concentration (Yb1200-30/ 250DC) was 13.9 dB/km at 1310 nm.
Spectral loss characteristics shown in Fig. 7 was obtained with tungsten halogen light source and OSA. The sample was fusion spliced to SMF pigtails at both ends. As described in Subsection 2.2, this caused considerable excitation of lossy cladding modes. Attenuation measured this way is significantly overestimated, in particular at short wavelengths. For example, total loss recorded at 1310 nm was 5.75 dB. Subtracting connector loss of 0.25 dB and twice the splice loss of 0.25 dB, we get net fiber loss of 5 dB and attenuation coefficient of 28.6 dB/km, more than twice the value measured with OTDR.

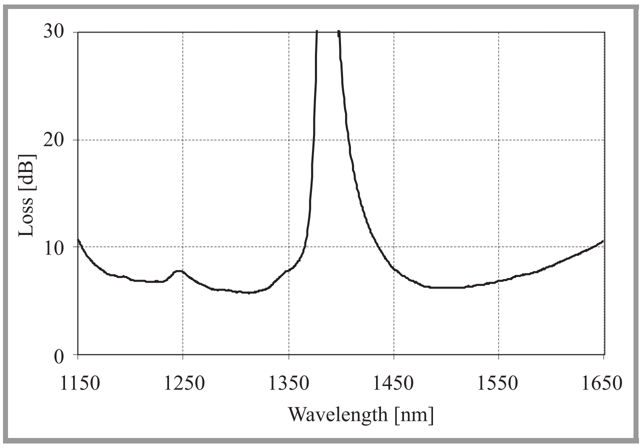


Fig. 7. Spectral attenuation of Liekki Yb1200-6/125DC fiber (length 175 m).

Accurate attenuation measurement using this setup would have required cutting off at least a 100 m long section of fiber to calibrate the OSA (see Fig. 4), making the measurement highly destructive.

Table 3
Polarization parameters of Yb1200-6/125DC fiber on 250 mm diameter spool – JME method

Parameter	Value
Fiber length [m]	200
Wavelength range [nm]	1480–1550
PMD (average DGD) [ps]	0.006
PMD coefficient [ps/km]	0.030
PDL (average) [dB]	0.03

Polarization properties were comparable to telecom SMF (see Table 3 and Fig. 8) with flat DGD spectrum. This indicates good control of fiber geometry and negligible forces exerted by coating.
Bending loss test, during which the fiber close to the end of 200 m length has been wound on set of mandrels, indicates low bending sensitivity (Fig. 9). Loss was measured with light emitting diode (LED) source and optical power meter. Loss instability was observed in several

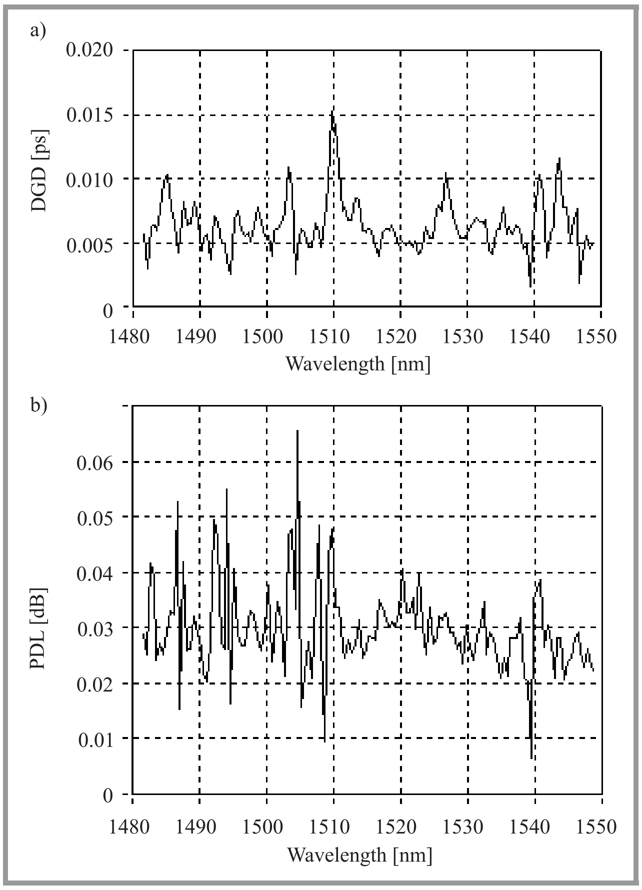


Fig. 8. DGD (a) and PDL (b) spectra (Liekki Yb1200-6/125DC, length 200 m).

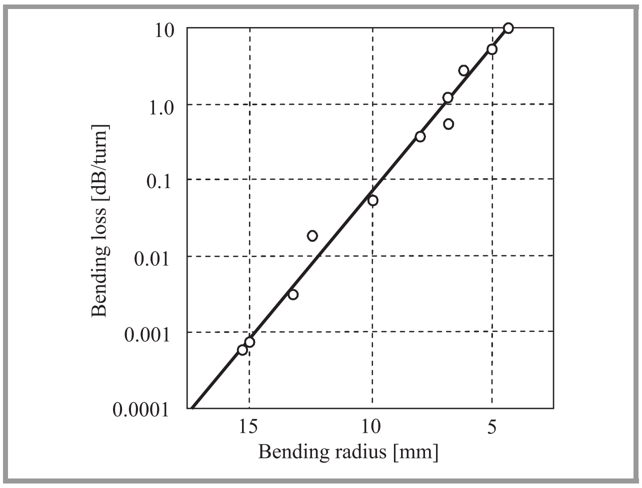


Fig. 9. Bending loss characteristics (Liekki Yb1200-6/125DC, $\lambda = 1300$ nm).

measurements, attributable to excitation of cladding modes, as the distance from mandrel to final splice before pigtail connected to power meter was only 1 m.

To investigate PMD created by bending (Table 4), a section of HDF approx. 2 m long was coiled on mandrels. This is a reasonable maximum length of fiber used in amplifier and usually enclosed in compact package.

Dramatic increase of PMD occurred when coil diameter was reduced below 45 mm; this bending limit is much stricter than dictated by loss. While the particular HDF is not intended for high-speed optical communication systems, erbium-doped fibers of similar design are likely to suffer from this problem, too.

Table 4
PMD introduced by bending (Yb1200-6/125DC, $\lambda = 1480$–1550 nm, total fiber length 200 m)

Bending radius [mm]	No. of turns	Length bent [m]	PMD [ps]	PDL [dB]
No bending	–	–	0.006	0.03
44	8	2.21	0.008	0.03
34	10	2.14	0.012	0.03
28	14	2.42	0.028	0.13
23	16	2.31	0.055	0.31
20	20	2.51	1.082	4.31
16	20	2.01	1.049	4.67

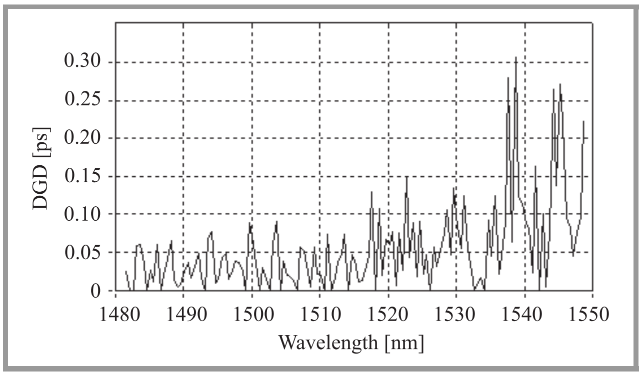


Fig. 10. DGD spectrum in Liekki Yb1200-6/125DC fiber bent at 23 mm radius

Such a great increase of PMD results from excitation of cladding mode(s), not strain-induced birefringence. This is confirmed by rapid increase of DGD with wavelength in threshold conditions (Fig. 10), caused by weaker guiding of fundamental mode at longer wavelengths.

Table 5
PMD introduced by crush (Yb1200-6/125DC, $\lambda = 1480$–1550 nm, total fiber length 200 m)

Crush force [N]	Pressure [N/m]	PMD [ps]	PDL [dB]
0	0	0.009	0.10
5	25	0.007	0.11
20	100	0.010	0.10
50	250	0.011	0.11
100	500	0.014	0.12

In another test, two parallel sections of PCF, each 100 mm long were crushed between flat steel plates. PMD variations were negligible until fiber coating was damaged when force reached 100 N (Table 5).

In summary, this type of HDF demonstrated excellent PMD performance, low bending loss and good quality of protective coating, but excitation of modes propagating in the inner cladding occurs easily when fiber is bent or spliced. Cladding modes propagate over long lengths.

3.2. Yb-doped fiber: Liekki Yb1200-10/125DC-PM

A 100 m sample of this fiber was used for COST-299 round-robin. Some results obtained at NIT will be presented here. This is a highly birefringent, double-clad, polarization-maintaining HDF fiber of PANDA design. Thanks to reasonable match of mode field diameter with telecom SMF (Corning SMF-28), there were no problems with splicing to SMF pigtails and related measurement artifacts. As the sample had to be delivered to other participants later in the same condition, mechanical experiments were ruled out.

Despite same level of doping, attenuation (Table 6 and Fig. 11) was higher than of Yb1200-6/125DC (Table 2 and Fig. 7). There is a strong OH^- absorption peak ($\approx$ 275 dB/km) and minimum of attenuation around 1300 nm. The tail of ytterbium absorption band is visible below 1150 nm.

Table 6

Attenuation of Yb1200-10/125DC-PM fiber – OTDR test

Wavelength [nm]	Attenuation [dB/km]
1310	20.6
1550	30.7

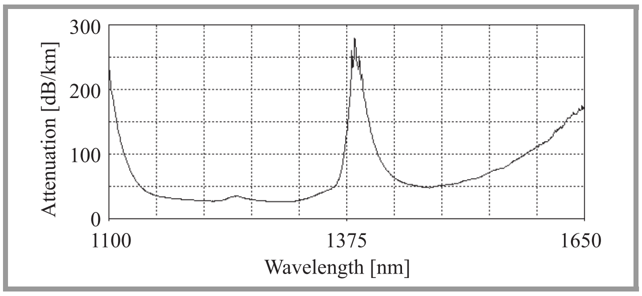


Fig. 11. Spectral attenuation of Liekki Yb1200-10/125DC-PM fiber.

Polarization properties were measured using both FA and JME methods, and for both directions of light propagation (see Table 7). Measured PMD values are in fairly good agreement with specifications [7] quoting birefringence $B \geq 1.4 \cdot 10^{-4}$, which corresponds to PMD of 467 ps/km. In a 100 m long sample, PMD was likely reduced by polarization mode mixing and true PMD coefficient may be higher.

Table 7

PMD of Yb1200-10/125DC-PM fiber on spool (fiber length 100 m)

Method	λ [nm]	Direction	PMD [ps]	PMD coefficient [ps/km]
JME	1500–1505	B-R	41.9	419
JME	1500–1510	R-B	42.3	423
JME	1500–1510	B-R	39.2	392
FA	1500–1505	R-B	42.5	425
FA	1500–1505	B-R	42.1	421

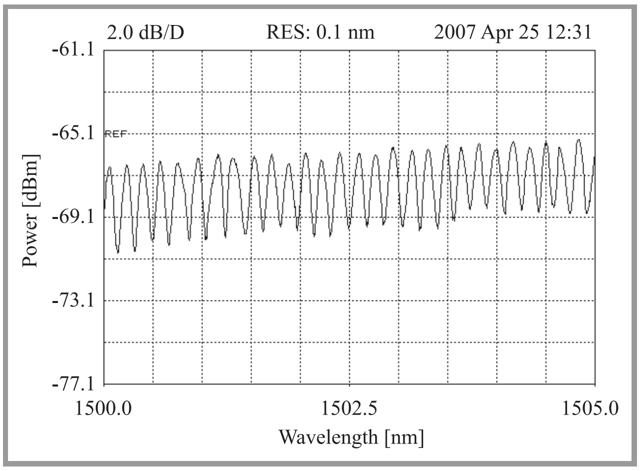


Fig. 12. Transmission spectrum in FA measurement (Liekki Yb1200-10/125DC-PM).

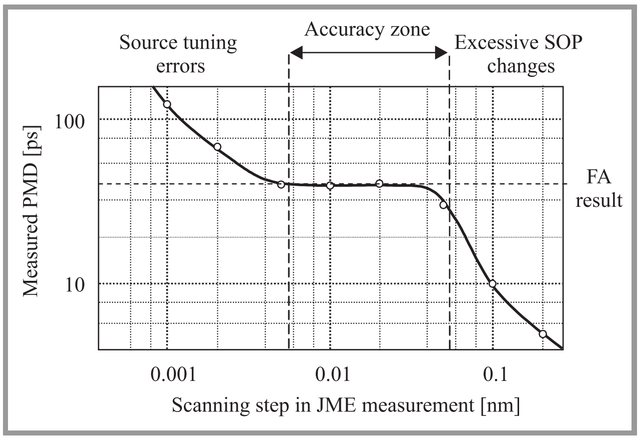


Fig. 13. Comparison of JME and FA measurements (Liekki Yb1200-10/125DC-PM, length 100 m).

While the FA method worked robustly, producing clear spectra (Fig. 12), PMD values delivered by JME analyzer depended on spectral scan step. Agreement with FA measurements was reached only in a certain range of settings (Fig. 13). This problem has been detected during tests of other fibers as well. It can be attributed to uncertainty of source tuning (HP 8168F) and possible "leaps" in detection of large shifts of state of polarization (SOP) when highly birefringent fiber is measured in large spectral intervals.

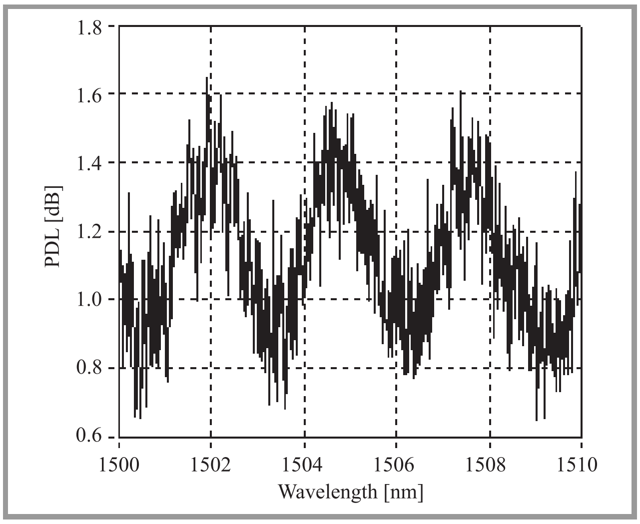


Fig. 14. PDL spectrum (Liekki Yb1200-10/125DC-PM, fiber length 100 m, direction R-B, scan step 0.01 nm).

Another interesting phenomenon was strong periodicity of PDL spectrum (Fig. 14).

3.3. PCF fiber: IPHT Jena 252b5

For this fiber, described in Subsection 2.2, PMD measurements (Table 8) and tests of influence of some external factors on PMD will be presented. A 17.1 m long sample was first tested during COST-299 round robin.

Table 8

PMD of IPHT Jena 252b5 holey fiber (sample length 17.1 m)

Method	λ [nm]	Direction	PMD [ps]	PMD coefficient [ps/km]
JME	1500–1520	B-R	21.5	1257
JME	1500–1520	R-B	21.6	1263
FA	1545–1555	R-B	23.2	1357

As the fiber exhibits strong birefringence, samples about 1 m long were used for further experiments.

Despite very small fiber core, suggesting purely single mode propagation, spectrum recorded during FA measurements indicated quite strong polarization mode coupling (Fig. 15), which may be a consequence of fiber geometry imperfections. Spectral distribution of DGD measured with JME analyzer was remarkably flat (Fig. 16), unlike results from tests of ordinary single mode fibers. Average PDL of the 17.1 m sample spliced to SMF pigtails was low: 0.30 dB.

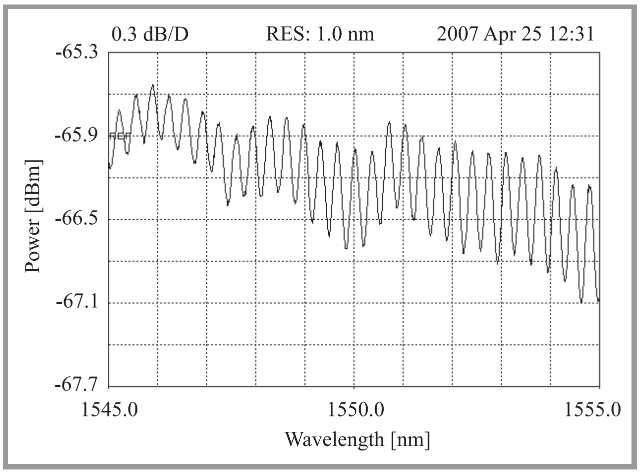


Fig. 15. Transmission spectrum in FA measurement (IPHT Jena 252b5, length 17.1 m).

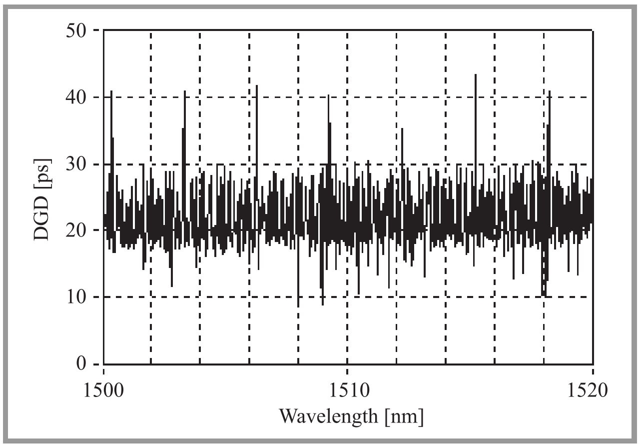


Fig. 16. DGD spectrum (IPHT Jena 252b5, fiber length 17.1 m, direction R-B, scan step 0.02 nm).

Rather surprisingly, PMD values obtained from JME and FA measurements have been in agreement for wide range of JME scan step between 0.002 and 0.1 nm (Fig. 17).

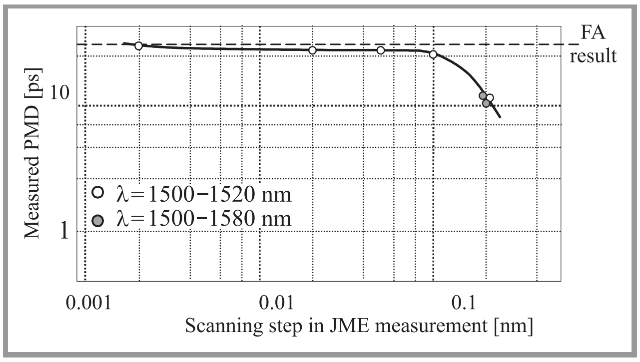


Fig. 17. Comparison of JME and FA measurements (IPHT Jena 252b5, length 17.1 m).

A 1.02 m long PCF was spliced to SMF pigtails, loosely placed on a flat plate and subjected to variable temperatures. This sample was taken from another length of PCF, which explains different PMD coefficient.

Table 9

IPHT Jena 252b5 – results of temperature cycling (fiber length 1.02 m, λ = 1480–1550 nm)

Temperature [°C]	PMD [ps]	PMD coefficient [ps/km]	PMD change [%]	PDL (average) [dB]
+20	1.104	1082	0.00	0.44
–20	1.100	1078	–0.36	0.46
0	1.102	1080	–0.18	0.42
+20	1.104	1082	0.00	0.45
+40	1.106	1084	+0.18	0.45
+60	1.107	1085	+0.27	0.52

Results (Table 9) indicate very low sensitivity of PMD to temperature; temperature coefficient was approx. $7.9 \cdot 10^{-5}$/K. For comparison, PANDA fibers exhibit temperature coefficients close to $9 \cdot 10^{-4}$/K.

Two other experiments on 0.82 m sample of PCF prepared in the same way were aimed at establishing effects of fiber bending and twist on its polarization properties. During the twist test, the fiber was kept straight at low tension. Splice protection sleeves were rotated against each other and twisted length of fiber was 0.80 m. Results are presented in Tables 10 and 11.

Table 10

IPHT Jena 252b5 – results of bending test (fiber length 0.82 m, λ = 1480–1550 nm)

Bending radius [mm]	No. of turns	Length bent [m]	PMD [ps]	PDL (average) [dB]
No bending	–	–	0.914	1.23
20	4	0.50	0.929	1.15
10	9	0.56	0.924	1.19
5	18	0.56	0.926	1.00

Table 11

IPHT Jena 252b5 – results of twist test (fiber length 0.82 m, λ = 1480–1550 nm)

No. of turns	Twist rate [rev/m]	PMD [ps]	Relative PMD	PDL (average) [dB]
0	0	0.914	1.000	1.23
16	20	0.874	0.956	0.93
32	40	0.749	0.819	0.81
48	60	0.663	0.725	0.86
64	80	0.590	0.646	1.03

Polarization properties of PCF were little affected by bending, and observed variations in PMD and PDL can be attributed mostly to measurement errors and problems with fiber handling.

Progressive twisting and resultant circular strain in fiber guiding area, however, resulted in steady, considerable reduction of PMD, although this effect was at least 100-fold weaker than in conventional single mode fibers investigated by the author [9].

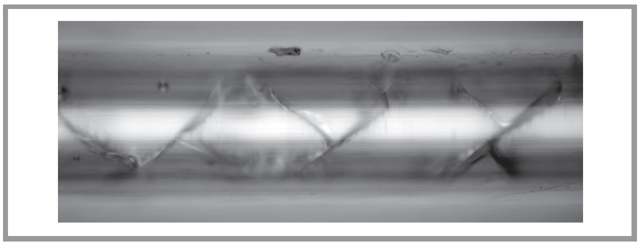


Fig. 18. Helical break in PCF cladding produced by excessive torsion.

When the twist rate reached 85 rev/m, the fiber was destroyed in a peculiar way: a section about 0.5 m long became soft and could elongate or bend with negligible stiffness; optical continuity was lost. Sample inspection with microscope revealed that a helical or double-helical break (Fig. 18) occurred in the cladding, but did not extend into the core.

4. Conclusions

Testing of PCF and HDF is difficult due to their unconventional optical and mechanical properties and difficulties with adopting commercial measuring instruments designed for testing of standardized telecom fibers. In particular, making a stable, low-loss coupling between specialty fibers of non-standard designs and core dimensions and test instruments requires additional work.

The PCF tested exhibited excellent stability of polarization parameters during mechanical tests and temperature cycling. This property and high PMD coefficient make it potentially useful for PMD etalons, e.g., for calibration of PMD analyzers or compliance testing of transmission systems.

Single mode HDFs, despite strong ytterbium doping, show PMD and bending performance comparable to standard SMF, proving good control of fiber geometry. However, dual cladding results in persistent propagation of cladding modes, causing severe problems during attenuation measurements.

Acknowledgements

The author is very grateful to our COST-299 partners, Kay Schuster and Jens Kobelke of IPHT Jena, Germany and Mircea Hotoleanu of Liekki – an nLight Company, Finland for supplying fiber samples.
Research work presented in this paper was financially supported by Polish Ministry of Science and Higher Education as special research project COST/39/2007.

References

[1] "Corning SMF-28e optical fiber product information", Corning Inc., PI1344, Jan. 2005.

[2] Y. Wang, H. Bartelt, S. Brueckner, J. Kobelke, M. Rothhardt, K. Mörl, W. Ecke, and R. Willsch, "Splicing Ge-doped photonic crystal fibers using commercial fusion splicer with default discharge parameters", *Opt. Expr.*, vol. 16, no. 10, pp. 7258–7263, 2008.

[3] B. Edvold and L. Gruner-Nielsen, "New technique for reducing the splice loss to dispersion compensating fiber", in *Proc. ECOC 1996 Conf.*, Oslo, Norway, 1996, vol. 2, pp. 245–248.

[4] B. Bourliaguet, C. Paré, F. Émond, A. Croteau, A. Proulx, and R. Vallée, "Microstructured fiber splicing", *Opt. Expr.*, vol. 11, no. 25, pp. 3412–3417, 2003.

[5] R. Thapa, K. Knabe, K. L. Corwin, and B. R. Washburn, "Arc fusion splicing of hollow-core photonic bandgap fibers for gas-filled fiber cells", *Opt. Expr.*, vol. 14, no. 21, pp. 9576–9583, 2006.

[6] "Yb1200-6/125DC single mode double-clad ytterbium-doped fiber", Liekki, Oct. 2007.

[7] "Yb1200-10/125DC-PM large mode area (LMA) double-clad ytterbium-doped fiber", Liekki, May 2007.

[8] "Transmission media characteristics – Optical fibre cables: Definitions and test methods for statistical and non-linear related attributes of single mode fibre and cable", ITU-T Rec. G.650.2 (01/2005).

[9] K. Borzycki, "Influence of temperature on PMD of tight buffered optical fibers and cables", *J. Telecommun. Inform. Technol.*, no. 1, pp. 56–66, 2008.

Krzysztof Borzycki was born in Warsaw, Poland, in 1959. He received the M.Sc. degree in electrical engineering from Warsaw University of Technology in 1982 and a Ph.D. degree in communications engineering from National Institute of Telecommunications (NIT), Warsaw in 2006. He has been with NIT since 1982, working on optical fiber fusion splicing, design of optical fiber transmission systems, measurement methods and test equipment for optical networks, as well as standardization and conformance testing of optical fiber cables, SDH equipment and DWDM systems. Other activities included being a lecturer and instructor in fiber optics and technical advisor to Polish fiber cable industry. He has also been with Ericsson AB research laboratories in Stockholm, Sweden, working on development and testing of DWDM systems for long-distance and metropolitan networks between 2001 and 2002 while on leave from NIT. He has been engaged in several European research projects since 2003, including Network of Excellence in Micro-Optics (NEMO) and COST Actions 270, 291 and 299. This included research on polarization mode dispersion in fibers and cables operating in extreme environments, and testing of specialty fibers. Another area of his work is use of optical fibers in electric power cables and networks. Dr. Borzycki is an author or co-author of 1 book, 2 Polish patents and 45 scientific papers in the field of optical fiber communications, optical fibers and measurements in optical networks, as well as one of "Journal of Telecommunications and Information Technology" editors.
e-mail: k.borzycki@itl.waw.pl
National Institute of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland

Paper

E-resources versus traditional teaching models

Alina Stasiecka, Ewa Stemposz, and Andrzej Jodłowski

Abstract—The paper presents the discussion about the e-resource structure and its influence on the resources' quality. The thesis we are taking into consideration is as follows: the conformance of e-resource structure with structures suggested by traditional teaching model/models has a strong influence on the quality of this e-resource. To achieve this, the most popular teaching models are analyzed and a proposal of the metamodel useful for e-resources construction is introduced.

Keywords— e-learning, e-resource, learning model.

1. Introduction

In publications dedicated to e-learning, various aspects concerning this form of education are considered. The quality of an e-resource is one of the key issues. According to researchers, it could make a significant impact on the effectiveness of e-teaching process as a whole. The main reason is the limited contact between a student and a teacher in this kind of teaching. Thus, the structure and the content of an e-resource should be well-thought-out to give more support to student during his/her self-study with the resource. Existing standards, like SCORM, IMS [1–3], define some requirements for the structure supporting e-resource construction. Unfortunately, they do not practically pay attention to the assessment of an e-resource quality in didactic and content-related aspects. According to the educators, *the quality of teaching process is strongly influenced by the degree of its conformance to the requirements defined among models existing in traditional teaching.* In the article, we present an approach to prove that the similar thesis is true also for e-teaching. We focus our attention on the quality of the resource from the didactic point of view, while its content-related aspect was left behind.

For introductory considerations, we chose the model of effective learning, discussed in details in our previous publications [8–13]. The model of effective learning, as the other ones used in traditional teaching, has a process nature. It is described by a sequence of particular stages which need to be applied in the appropriate order. Thus, the first step is an adaptation to e-learning needs. It requires the transformation of the teaching process into e-resource structure where the structure' elements preserve both the order of process stages and stages' time proportions. To generalize the results, further examinations were carried out on others traditional teaching models.

The paper is organized as follows: Section 2 describes the already achieved results. Section 3 presents the characteristics of the most popular groups of traditional didactic models. Section 4 introduces the proposal of a teaching metamodel that constitutes the basis for e-resources construction. Section 5 concerns the discussion on the correspondence between e-resource structure and the structure defined by the metamodel, and the dependency between e-resource structure and its quality. Section 6 contains the conclusions and plans for future works.

2. E-resource quality versus the model of effective learning

In this section we recall our previous research concerning the following thesis: *the quality of e-resource is conformant to the correspondence between its structure and the ones suggested by the existing traditional teaching models.* The research was done on the basis of the model of effective learning [5, 8, 13, 15].

According to the model, a teaching resource should be constructed hierarchically with two levels of hierarchy. The first level includes four elements, such as:

1. Introduction,
2. Main content,
3. Summary,
4. Evaluation.

The second level should contain the elements which are nested in the appropriate element of the higher level (see Fig. 1).

Additionally, the structure should support some limitations put on it by the model. For the first level of hierarchy: all elements should be present in the resource, they should be kept in the right order, and they should be kept in the appropriate proportions (10%, 65%, 15%, 10% of the whole resource). For the second level: as previously, the presence of all the elements and their order are still required, whereas there are no limitations on the elements proportions. We assumed that a resource would be conformant to the model of effective learning if its structure is organized according to the requirements mentioned above.

To prove the thesis, the examination of 56 virtual e-resources was conducted. E-resources were acquired by the instructors and the students of Warsaw technical universities. The space of features (measures) was created on the basis of the resource structure conformance to the structure defined by the model.

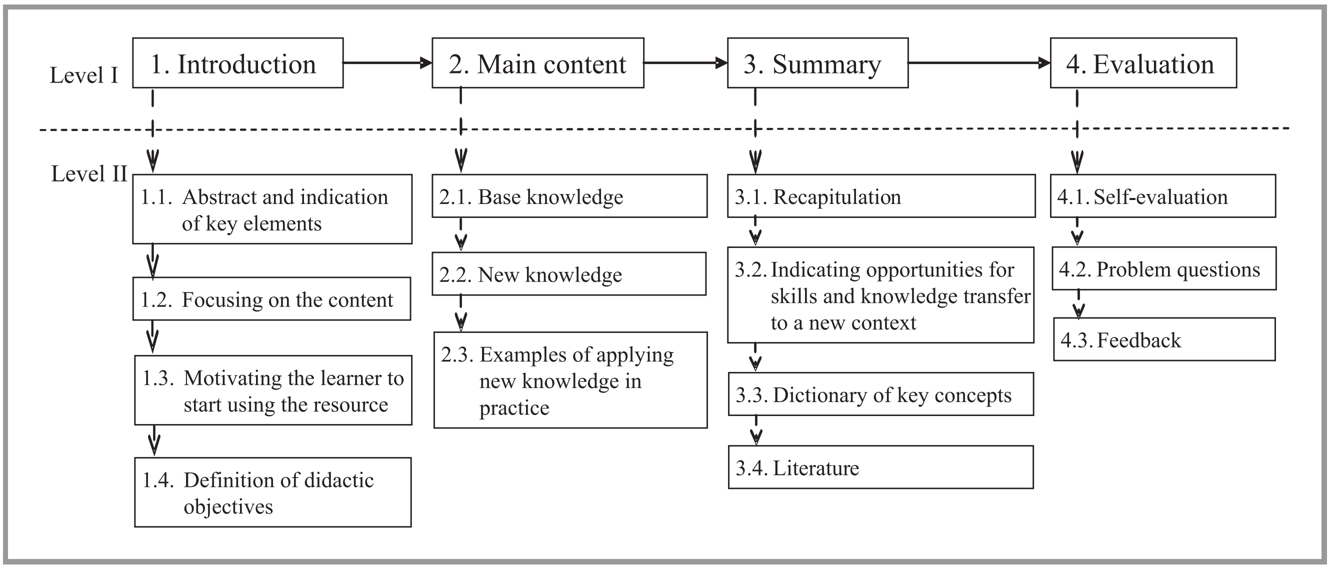


Fig. 1. The structure of a resource conformant to the model of effective learning.

For each measure, we introduced the following notation:

position_in_resource. suffix

where:

position_in_resource – denotes the nesting path for a partial element connected with considered measure. For example, if position_in_resource is 2 it means that the measure describes the second element on the first hierarchy level (2. Main content), while 2.3. marks the third partial element of the second level (2.3. Examples of applying new knowledge in practice).

suffix – defines the kind of measure, as following:

- p – means the presence of a measure-connected-element in e-resource. The measure takes an integer value from the interval [0,1], where 1 means that the element is present within the resource, while 0 – means its lack. For example, for the element 2.3. Examples of applying new knowledge in practice, the measure 2.3.p denotes the presence of this element within the resource.
- q – means the quality assessment of the considered element given by a respondent; the measure takes an integer value from the interval [0,5]. For example, for the element 2.3. Examples of applying new knowledge in practice, the measure 2.3.q denotes its quality.
- t – defines the ratio (in percentage terms) of the estimated time devoted to work with the considered element to the time devoted to work with the resource as a whole. Traditional educators were estimating the proportions of partial elements through counting the pages of materials. Therefore, it was necessary to adapt the semantics of the measure to the complex virtual environment, where e-resources have a non-linear nature and can contain multimedia elements, etc. The measure takes the value from the interval [0, 100]. For example, if the value 1.t is 7 [%], it means that the element 1. Introduction takes 7% of the whole resource.

For a resource conformant to the model of effective learning, the F_{mefl} and F'_{mefl} measures' spaces are defined as follows:

$$F_{mefl} = \{1.p,\ 1.q,\ 1.t, \ldots, 4.p,\ 4.q,\ 4.t,\ 1.1.p,\ 1.1.q, \ldots, 1.4.p,\ 1.4.q,\ 2.1.p,\ 2.1.q, \ldots, 4.3.p, 4.3.q\},$$

$$F'_{mefl} = \{1.p,\ 1.q,\ \ldots,\ 4.p,\ 4.q,\ 1.1.p,\ 1.1.q,\ \ldots,\ 1.4.p,\ 1.4.q,\ 2.1.p,\ 2.1.q,\ \ldots,\ 4.3.p,\ 4.3.q\},$$

where the F'_{mefl} is a subset of F_{mefl} with the t-measures excluded.

In the next step, we introduced a virtual ideal e-resource to compare it against the whole examined population. The ideal resource is the one containing all required partial elements, where elements are placed in the correct order and, at the first hierarchy level, their appropriate proportions are kept.

To provide the multidimensional data analysis, we exploited ARs' tables, one of the statistical program GradeStat tools [4].

The concentration index AR allows the evaluation of a distance between a considered e-resource and the ideal one. The smaller AR values correspond to greater similarity with the model. The values of concentration index belong to the interval [0,1].

Figure 2 contains the AR chart constructed on the basis of the F'_{mefl} set and the 56 e-resources population while Fig. 3 presents the results for F_{mefl} set and 37 e-resources. In the second case, only those resources are taken into considerations for which the values of t-measures were given by the respondents.

On the basis of the presented results, one can notice that e-resources with the structure more conformant to the model of the effective learning, achieved better marks from the respondents than the ones with the lower conformance.

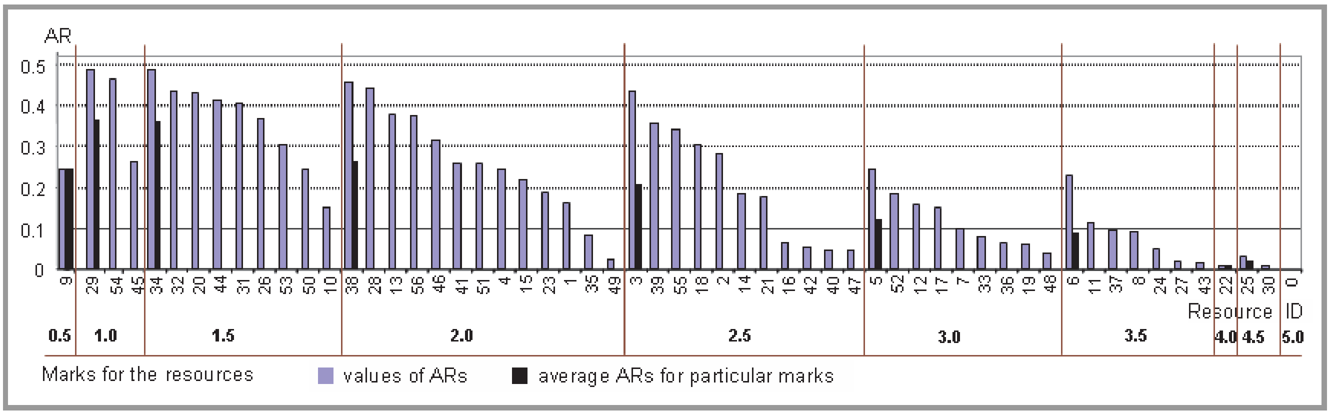


Fig. 2. ARs for F'_{mefl} and the population of 56 e-resources.

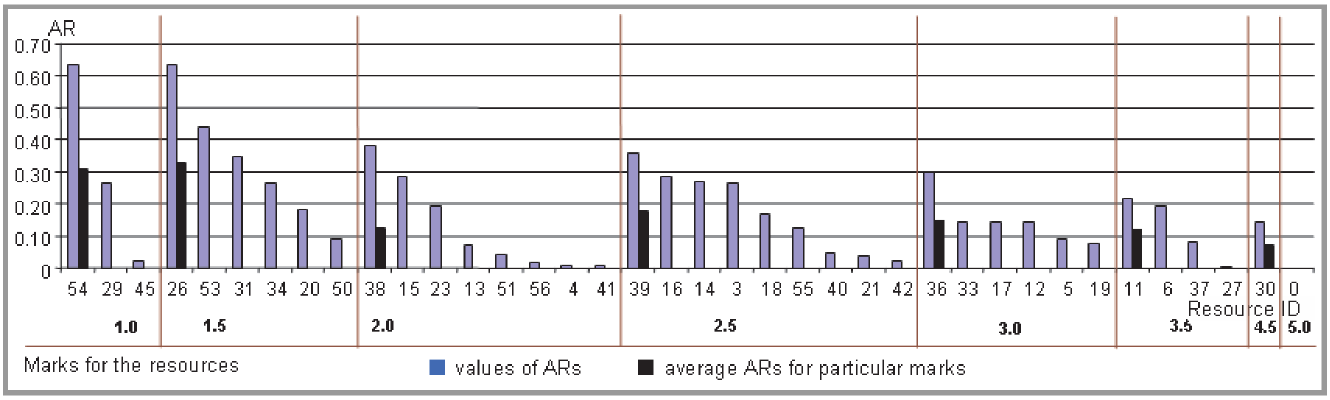


Fig. 3. ARs for F_{mefl} and the population of 37 e-resources.

To generalize the thesis formulated in Section 2 onto any traditional teaching model, further research was done. We performed an analysis of several other models to identify their common features and on the basis of that, to construct both the metamodel of teaching and the new space of measures useful for the thesis verification.

3. Other teaching models used in traditional teaching

A model of teaching/learning (a didactic model also called *a strategy of teaching*) defines: content, methods of learning and learner's didactic environment [7].

There is a variety of teaching models/group of models which can supplement each other [6], so it is possible (and recommended) to combine different methods and styles into one didactic process. A teacher should choose an appropriate strategy of teaching depending on the goals that he is going to achieve. In this section, we present a more detailed analysis of some of the models considered by us as the most useful for e-learning:

Process-recognition models focus on improving the learner's mental abilities to support him/her in acquiring information, creating notion and in promoting creative thinking. That group includes: the induction model, notion creation model, synectics model, mnemonic model, model of effective learning (discussed in Section 2).

Behavioral models – the ones which base on the behavioral theory, where a person (treated as "a black box") is considered as a kind of"self- improving communication system" which modifies its behavior in response of reverse information. Among them: the social teaching, program teaching, simulations, and the direct teaching models could be found.

Social models support learning of cooperation methods, stimulate activity, facilitate the usage of other students' work, and examine social relations. That group includes: teaching through joint research, role-playing models, etc. According to us, social models, because of their specific features, are not very useful for e-learning needs.

Personal development models pay attention to the learner's internal development, integration of different aspects of learner's personality (emotions, intellect). They are focused on the stimulation of development with the active teacher's support on each step of education. Therefore, they are too difficult to use in case of self-work and as a result not useful for e-learning. The non-directive learning would be a typical representative of this group.

Induction model we chose for more detailed discussion. (During the research, the other models suitable for e-learning were analyzed in a similar way.) The didactic goal of the induction model is, apart from gaining new knowledge, acquiring the ability to form notion categories and to use them in the proper context.

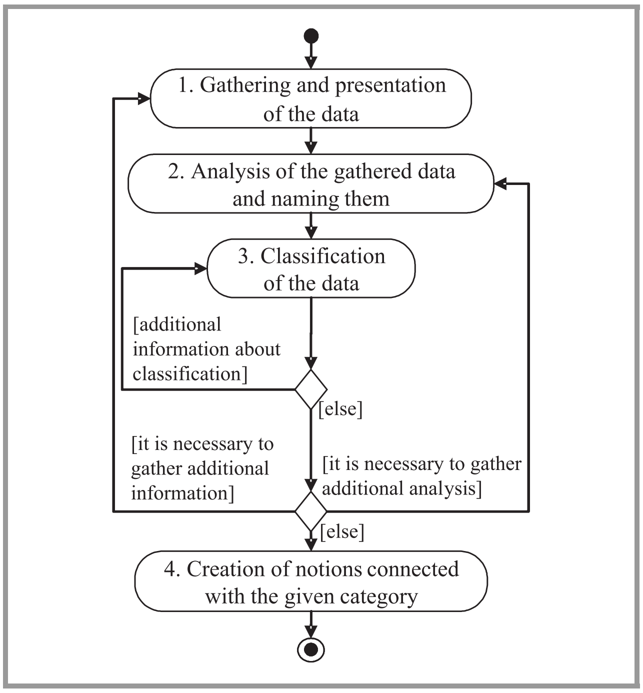


Fig. 4. Induction model.

In the induction model, the following stages can be distinguished (see Fig. 4):

1. Gathering and presentation of the data

   The teacher outlines the general area where the data should come from, for example natural medicine. The learners gather the data related to the given topic. It is possible that in the further stages it will be necessary to add/remove the data. When process is coming back to the stage 1, the gathered data are reorganized.

2. Analysis of the gathered data and naming them

   The data gathered in stage 1 are mostly incidental and chaotic. Further analysis should be performed in order to: assign the names to them and describe them using characteristic features. As a result, it is possible to define appropriate ordering of data and facilitate making use of them.

3. Classification of the data

   Classification of the data is necessary to exploit them in the following steps. The classification stage is performed several times. The first classification is very general. After that, it is possible to make more detailed classification. It happens quite often that there is a need to gather additional information (back to stage 1) or to carry out additional analysis (back to stage 2). Then, the renewed classification of data is executed.

4. Creation of notions connected with the given category

   After the data classification the learners are able to recognize the features of notions and to assign notions to the appropriate category. They also acquired the ability to formulate notions which are conformant to the given category.

4. Metamodel of learning

As a result of the analysis of the learning models discussed in Section 3, we noticed that some of them are effective only in the case of traditional teaching. It could be difficult to use them in e-learning, especially due to lack of frequent interaction with the teacher, what is characteristic for this type of learning. In the further research we skipped some of them. While considering the other models, it was possible: to extract some common features and to create the metamodel of teaching which could be useful in constructing e-resources with the structure corresponding to the ones used in traditional teaching. The structure of the metamodel is presented in Fig. 5.

To keep the picture clear, we avoided any comments and constraints. They were put below together with explanations concerning the values of attributes, the conditions of their optionality and derivation.

The Learning Model class

The derived attribute *catalogue name* is defined as following: the name of the learning models category + the name of a model's basic version (for which the number of version equals 0) + the name of the model's version + the nesting level of the version.

The model is made by the constructor(s) and may be extended by the reconstructor(s) – the person(s) working on the next version(s) of the model.

Interpretation of the attribute's value *constructors_reconstructors* depends on the version number. If it is the model's basic version, then the values concern the constructors of the model. If it is the subsequent model's version, then the values describe the reconstructors.

The attribute *description of changes* denotes short characteristics of the most important changes made to the model in comparison with the previous version – as a text.

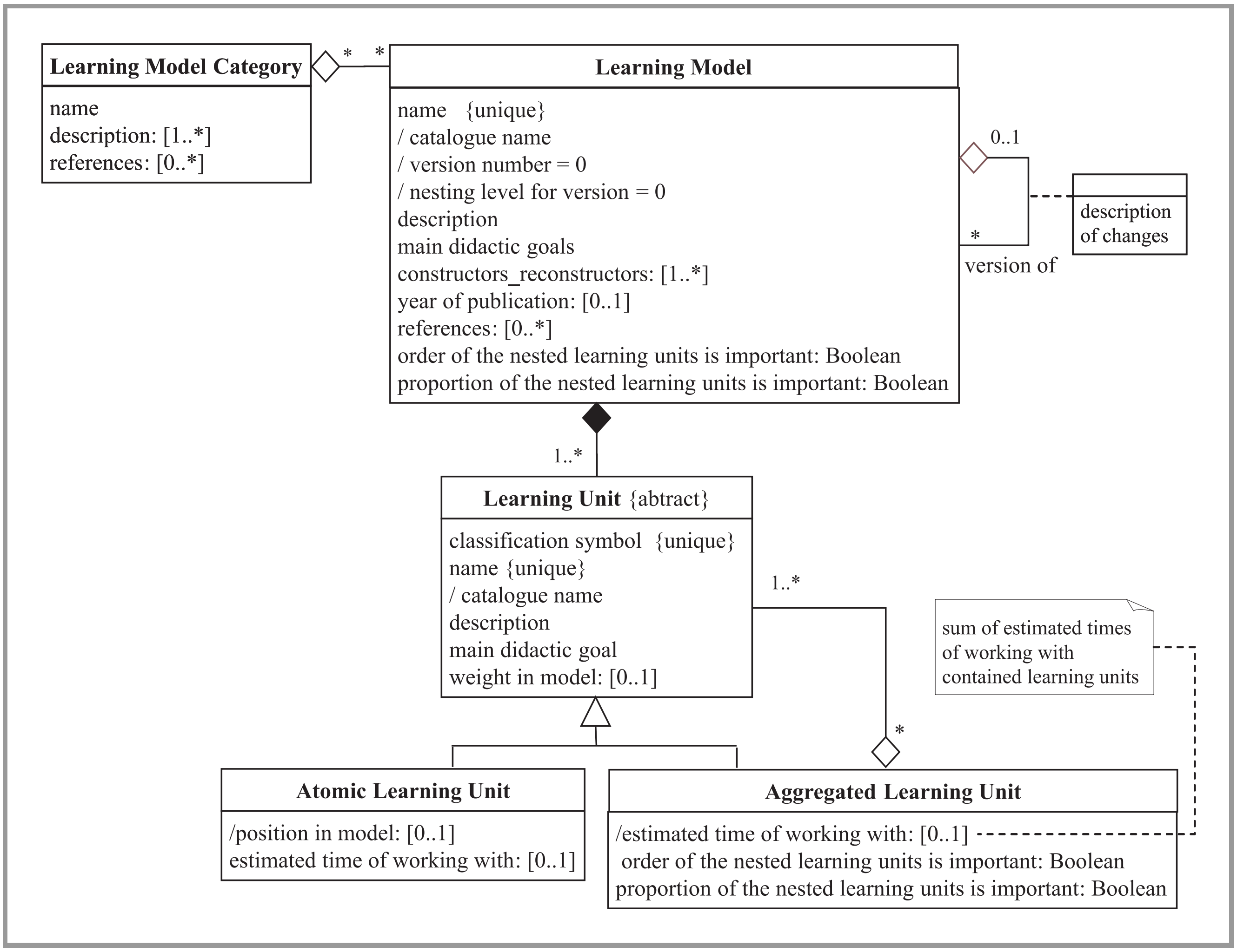


Fig. 5. The structure of a metamodel.

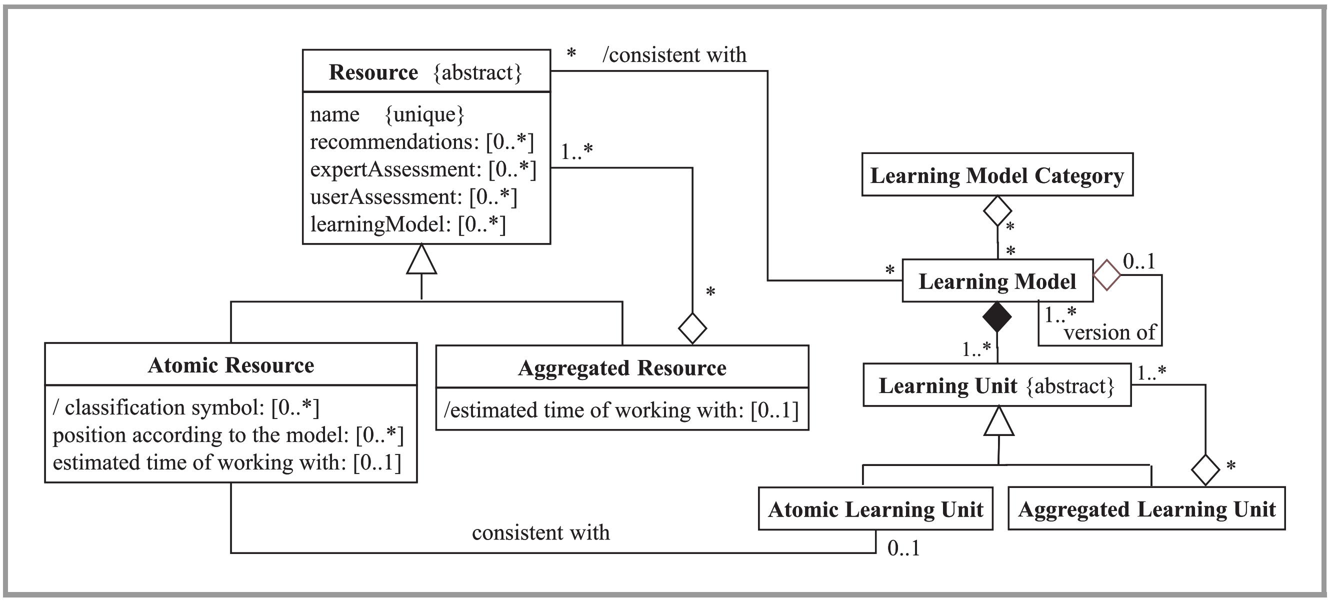


Fig. 6. The structure of a resource.

The Learning Unit class

The derived attribute *catalogue name* is defined as follows: the name of the model's particular version + the name of the learning unit.
The value of the attribute *classification symbol* means if the mutual order of the nested units is important. If it is, then the attribute's value is stated as the unit position conformant to the relevant unit defined in the model, e.g., "1.1", "1.1.1". In the opposite case, letters are used, e.g., "1.A", "1.1.B".
According to the metamodel definition, we proposed the e-resource structure conformant to the metamodel (see Fig. 6). To keep the picture clear, the Learning Model and Learning Unit classes were presented in the simplified version (without attributes, methods and associations which are not important from this point of view).

5. E-resource quality versus traditional teaching models

To generalize the results of the research, which were briefly discussed in Section 2, a few next steps were taken. At the beginning we made an attempt to define the notion of *the conformance of e-resource with any traditional teaching model*. We assumed that a resource is conformant to the chosen model, if: it contains all elements required by the model, and nested elements are both properly ordered and their proportions are kept.
To prove the thesis, we did research on the quality for the same 56-element population of e-resources (see Section 2).

The construction of the new space of measures was based on the metamodel of teaching. The new measures were only related to the resource's structure. In opposition to the model of effective learning, the measure related to the quality assessment was excluded.
As before (see Section 2), each measure is described by the following expression:

position_in_resource. suffix

where:

position_in_resource – has the same meaning as for the measures' space of the model of effective learning,

suffix – denotes the kind of measure, where:

p – has the same meaning as before,

t – has the same meaning as before, but this measure may be defined on any nesting level in the model.

To this measures' space, there was introduced a new measure – o (order). The measure o concerns every aggregated element (having nested elements). It takes the values from the interval [0, 100], which is given in per cent. The measure value = 100% means correct order of the nested elements, each value $< 100\%$ points e-resource with incorrect structure.

The level of partial elements' order preservation may de defined as for example: the ratio of lr to mlr, where lr means the number of movements which need to be done for the resource to achieve the required order, and mlr defines the maximal number of movements, assuming that the order is a total opposite of the advised one.

There are two typical cases:

1. If an e-resource is an aggregated element (e.g., a complete course), then 100% of conformance to the model/models means that a resource contains all the advised partial elements, which are ordered correctly and kept in the appropriate proportions.

2. If an e-resource is an atomic element, then we assume it is of 100% conformance with the considered model/models.

On the basis of above considerations, the space of measures useful for the quality examination was constructed. In the next, we presented two examples of the measures' space for the complete course resources satisfying requirements defined by two following traditional teaching models:

– model of effective learning:

$$F_{mefl'} = \{0.o,\ 1.p,\ 1.t,\ 1.o,\ 2.p,\ 2.t,\ 2.o,\ \ldots,\ 4.p,\ 4.t,\ 4.o,\ 1.1.p,\ 1.2.p,\ \ldots,\ 4.3.p\},$$

– induction model:

$$F_{mi'} = \{0.o,\ 1.p,\ 2.p,\ 3.p,\ 4.p\},$$

where level 0 means the measure connected with the resource treated as a whole.
It is easy to observe that in the case of the induction model there are no measures with suffix t, because no suggestions of the appropriate proportions of the elements we found in the literature.
To verify the thesis considered in the paper, the following research of the 56-element population was carried out. The research was done for the new measures' space, constructed for the model of effective learning on the basis of the defined metamodel. As previously (see Section 2), the statistical program GradeStat was used.
Figures 7 and 8 present two AR charts for the new measures' space. Figure 7 concerns the population of 56 e-resources and the measures' space:

$$F_{mefl''} = \{0.o,\ 1.p,\ 1.o,\ 2.p,\ 2.o,\ \ldots,\ 4.p,\ 4.o,\ 1.1.p,\ 1.2.p,\ \ldots,\ 4.3.p\},$$

where t measure was excluded.

Figure 8 contains the chart for the population of 37 e-resources and $F_{mefl'}$ measures' space.
While analyzing the trends in the charts, it is easy to notice that the e-resources highly conformant to the model of effective learning (low AR values) got better marks from respondents than the resources with more differences between their structures and the structure required by the model.

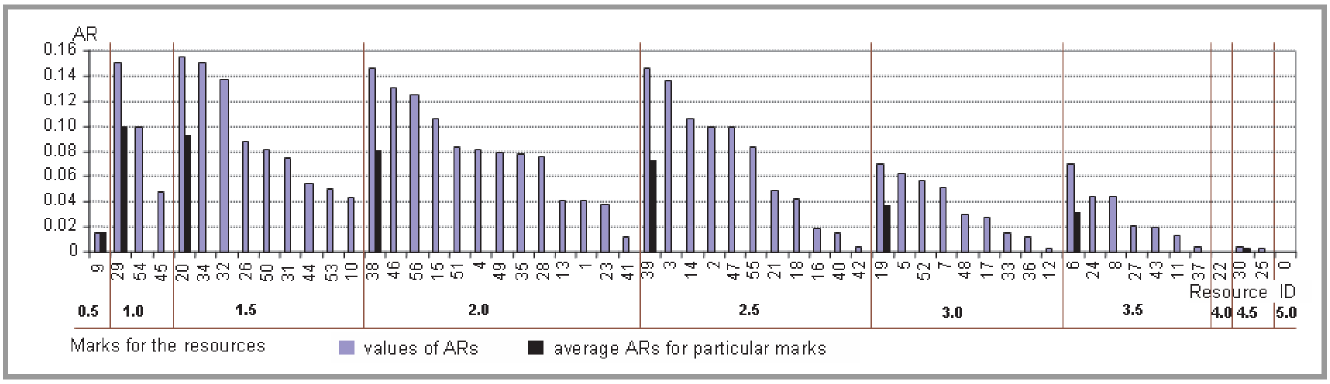


Fig. 7. ARs for $F_{mefl''}$ and the population of 56 e-resources.

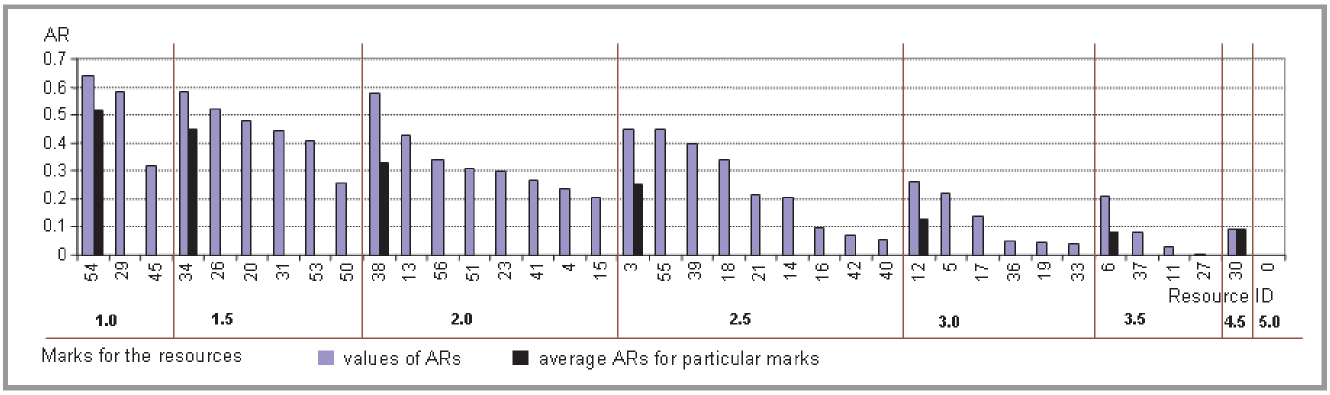


Fig. 8. ARs for $F_{mefl'}$ and the population of 37 e-resources.

Summarizing, the analysis of the data allowed the positive verification of the thesis: *The conformance of an e-resource structure to at least one traditional teaching model has an influence on the quality of the resource.*

6. Conclusions and future research

In the paper we presented the research concerning the influence of the e-resource structure conformance with the structures required by traditional teaching models on e-resource quality, from a didactic point of view.
As a result, we presented:

– the metamodel for traditional teaching models,

– the e-resource structure which is conformant to the structure defined by the metamodel.

We will continue our research to enhance the metadata currently existing in e-learning standards. This concerns the elements/categories which may have influence on the resource quality, both in didactic and non-didactic aspects (e.g., related to its potential to reuse).

References

[1] ADLNET, 2004. SCORM 2004 standard, http://www.adlnet.org/

[2] IMS standard, 2004, http://www.imsglobal.org

[3] LOM standard, 2002, http://ltsc.ieee.org/wg12/

[4] "Program for grade data analysis", GradeStat, 2007, http://gradestat.ipipan.waw.pl/

[5] S. M. Allesi and S. R. Trollip, *Multimedia for Learning: Methods and Development*. Needham Heights: Allyn and Bacon, 2001.

[6] B. Joyce, E. Calhoun, and D. Hopkins, *Przykłady modeli uczenia się i nauczania* (Models of learning – tools for teaching). Warsaw: WSiP, 1999 (transl. – in Polish).

[7] W. Okoń, *Nowy słownik pedagogiczny*. Warsaw: Wyd. Akademickie Żak, 2004 (in Polish).

[8] A. Stasiecka, W. Bednaruk, W. Dabrowski, and E. Stemposz, "Didactic aspects of e-learning resources", in *Proc. Int. Conf. Interact. Comput.-Aided Learn.*, Villach, Austria, 2004.

[9] A. Stasiecka, E. Stemposz, and W. Dabrowski, "Didactic aspects influence on quality of e-learning resources", in *Proc. WSEAS Int. Conf. Circ. Syst. Commun. Comput.*, Athens, Greece, 2005 (*WSEAS Trans. Inform. Sci. Appl.*, iss. 7, vol. 2, pp. 1002–1008, 2005).

[10] A. Stasiecka, J. Plodzien, and E. Stemposz, "Towards improving the quality of e-learning resources", in *Proc. Metadata Semant. Res. MTSR 2005*, http://www.metadata-semantics.org

[11] J. Plodzien, E. Stemposz, and A. Stasiecka, "An approach to the quality and reusability of metadata specification for e-learning objects", *Int. J. Dig. Inform. Res. Use*, vol. 30, no. 3, pp. 238–251, 2006.

[12] E. Stemposz, A. Stasiecka, and J. Plodzien, "Identyfikacja zbioru miar dla przybliżonego szacowania jakości zasobów e-learningowych", in *Proc. VI Conf. Virtual Univ. VU 2006*, Warsaw, Poland, 2006 (in Polish).

[13] E. Stemposz, A. Stasiecka, and A. Jodlowski, "The proposal of metadata for defining the quality of e-learning", in *Proc. Conf. ENMA 2007 Eng. Math.*, Bilbao, Spain, 2007.

[14] E. Stemposz and A. Stasiecka, "Determining a set of measures for quality estimation of e-resources conformant to the model/models defined in traditional education", in *Proc. 6th WSEAS Int. Conf. E-ACTIVITIES'07*, Puerto de la Cruz, Spain, 2007.

[15] A. Stasiecka, J. Plodzien, and E. Stemposz, "Measures for estimating the quality of e-learning materials in the didactic aspect", in *Proc. Conf. WEBIST 2006*, Lisbon, Portugal, 2006.

Alina Stasiecka was born in Warsaw, Poland. She finished Electronics and Information Technology Faculty of the Warsaw University of Technology. She was working in the Institute of Computer Science of the Polish Academy of Science (years 1986–2001). Since 2001 she has been employed at Polish Japanese Institute of Information Technology, Warsaw, where she has been engaged in software engineering and e-learning.

e-mail: alas@pjwstk.edu.pl
Polish Japanese Institute of Information Technology
Koszykowa st 86
02-008 Warsaw, Poland

Ewa Stemposz was born in Wasilków/Białystok, Poland. She graduated master degree at the Faculty of Electronics, Technical University of Warsaw. She has been employed at the Institute of Computer Science of the Polish Academy of Sciences (from 1988) and the Polish Japanese Institute of Information Technology (from 1994), where she has been engaged in computer graphics, computer vision, data basis, software engineering, object analysis, project management, and e-learning.

e-mail: ewag@ipipan.waw.pl
Institute of Computer Science
Polish Academy of Sciences
J. K. Ordona st 21
01-237 Warsaw, Poland

e-mail: ewag@pjwstk.edu.pl
Polish Japanese Institute of Information Technology
Koszykowa st 86
02-208 Warsaw, Poland

Andrzej Jodłowski is a Lecturer at Higher School of Entrepreneurship and Social Sciences, Poland. He received his M.Sc. degree from the Warsaw University of Technology in 1995, in the area of applied computer sciences. He received his Ph.D. from the Institute of Computer Science of the Polish Academy of Sciences in 2003, in the area of object-oriented database technologies. His research interests concern object-oriented databases, software engineering, and e-learning.

e-mail: andrzejj@pjwstk.edu.pl
Higher School of Entrepreneurship
and Social Sciences
Armii Krajowej st 13
05-400 Otwock, Poland

Polish Japanese Institute of Information Technology
Koszykowa st 86
02-008 Warsaw, Poland