

USING ENSEMBLE AND LEARNING TECHNIQUES TOWARDS EXTENDING THE KNOWLEDGE DISCOVERY PIPELINE

Yu-N Cheah, Sakthiaseelan Karthigasoo, Selvakumar Manickam
School of Computer Sciences, Universiti Sains Malaysia, 11800 USM Penang, Penang, Malaysia
Email: seelan@cs.usm.my, yncheah@cs.usm.my, selva@cs.usm.my

Keywords: Knowledge discovery, Clustering ensemble, Neural network ensemble, Discretization, Rough set analysis

Abstract: Knowledge discovery presents itself as a very useful technique to transform enterprise data into actionable knowledge. However, their effectiveness is limited in view that it is difficult to develop a knowledge discovery pipeline that is suited for all types of datasets. Moreover, it is difficult to select the best possible algorithm for each stage of the pipeline. In this paper, we define (a) a novel clustering ensemble algorithm based on self-organizing maps to automate the annotation of un-annotated medical datasets; (b) a data discretization algorithm based on Boolean Reasoning to discretize continuous data values; (c) a rule filtering mechanism; and (d) to extend the regular knowledge discovery process by including a learning mechanism based on neural network ensembles to produce a neural knowledge base for decision support. We believe that this would result in a decision support system that is tolerant towards ambiguous queries, e.g. with incomplete inputs. We also believe that the boosting and aggregating features of ensemble techniques would help to compensate for any shortcomings in some stages of the pipeline. Ultimately, we combine these efforts to produce an extended knowledge discovery pipeline.

1 INTRODUCTION

The generation of a huge amount of data by an enterprise is of great concern to decision makers. This problem is compounded by the many environmental challenges that an enterprise faces in the effort to produce better products and services. It is highly crucial to know what goes on in its business transactions both internally and externally and to examine the heart of an enterprise's transactions, that is its data, and to transform it into actionable knowledge through the process of knowledge discovery.

Knowledge discovery is a series of processes, which can be likened to a pipeline to find hidden but potentially useful information and patterns in data (Fayyad et al., 1996). This series of processes involves preparation on the data, the application of data mining algorithms on the data and finally the interpretation and/or visualization of the data mining results. The number of ways that a pipeline can be developed is perhaps limitless in view that there are many ways in which data preparation, data mining and interpretation or visualization can be achieved.

Current knowledge discovery pipelines have been proven effective to a certain extent in discovering hidden knowledge from various

datasets. However, their effectiveness is limited as it is difficult to develop a pipeline that is suited for all types of datasets. Moreover, it is difficult to select the best possible algorithm for each stage of the pipeline. It is challenging to develop an approach that would minimize the impact of a sub-optimal choice of algorithm at each stage of the pipeline.

Therefore, in this paper, we aim to define (a) a novel clustering ensemble algorithm based on self-organizing maps (SOM) to automate the annotation of un-annotated medical datasets; (b) a data discretization algorithm based on Boolean Reasoning to discretize continuous data values; (c) a rule filtering mechanism employing a Rule Quality Function; and (d) to extend the regular knowledge discovery process by including a learning mechanism based on neural network ensembles (NNE) to produce a neural knowledge base for decision support. We believe it is advantageous to produce a knowledge base that is trained from decision rules to develop a decision support system that is tolerant towards ambiguous queries, e.g. with incomplete inputs. We also believe that the boosting and aggregating features of ensemble techniques would help to compensate for any shortcomings in some stages of the pipeline. Ultimately, we combine

these efforts to produce an extended knowledge discovery pipeline.

1.1 Ensembling Techniques

Ensembling is a technique that harnesses the capabilities of a predetermined number of algorithms or processes that perform the same task to obtain an improved result.

Ensembling involves two main steps. The first is the execution of the individual algorithm. Two popular approaches for this are boosting (Freund and Schapire, 1995) and bagging (bootstrap aggregation) (Breiman, 1996). In boosting, the output of a particular individual algorithm becomes the input for another algorithm. The outputs are boosted through re-sampling or re-weighting until all the algorithms have had a hand in processing the input. In bagging, several subsets are derived from the original input and each one is fed to a different algorithm to be processed separately. The second step involves the combination of the outputs to produce a single consolidated output, i.e. as if only a single instance of the algorithm was used. Techniques such as voting and averaging are popularly used for this purpose.

An example of an ensembling technique is the NNE (Hansen and Salamon, 1990) where the learning capabilities of a predetermined number of neural networks is utilized to obtain improved generalizations or predictions (Zhou, et al., 2003). Another example of the ensembling technique is in the area of clustering (Yang and Kamel, 2003).

1.2 Knowledge Discovery Pipelines

An example in developing a knowledge discovery pipeline for symbolic rules extraction from un-annotated datasets is by Abidi and Hoe (2002) which applies rough set analysis. This pipeline, or workbench as they have called it, includes steps to pre-process and cluster un-annotated datasets resulting in annotated versions of the original datasets. The workbench then proceeds to discretize the data and generate rules. These rules are finally filtered.

For clustering, the workbench employs the K-Means algorithm while the Chi Squared and Entropy-MDL algorithms were used to discretize the

annotated version of the data. Rough set analysis was then used to compute the reducts and generate symbolic rules. Here, the reducts were derived using genetic algorithm. The rules were then filtered using a rule quality index computation.

Another work on a knowledge discovery pipeline involves the discretization of numerical attributes for machine learning (Risvik, 1997). This pipeline includes steps to pre-process and then discretize annotated datasets using discretization algorithms such as Chi Squared, Entropy, Naïve and Orthogonal Hyperlanes. Rough set analysis was used to generate rules from the discretized data. Reducts is generated using Johnson's Algorithm (1974) and followed by a rule generation process.

From our observation, the stages involved in knowledge discovery are more or less standardised and normally ends with a filtering stage. It appears that much of the research in knowledge discovery pipelines involves the exploration of various algorithms that can be applied at each stage of the pipeline. We believe that existing pipeline construction can be extended further by including a learning stage and that the algorithms can be made more effective by employing ensemble techniques, i.e. by being made to 'work harder' to compensate for any shortcomings as mentioned earlier.

2 THE EXTENDED KNOWLEDGE DISCOVERY PIPELINE

Our extended knowledge discovery pipeline consists of six stages (see Figure 1): data preparation, clustering, discretising, rough set analysis, rule filtering and learning.

2.1 Stage I: Data Preparation

Data preparation ensures the cleanliness or completeness of the dataset. Various strategies are currently employed to remove redundant data, missing values and other errors. For our purposes, we use the mean/mode fill technique to address the issues of missing values. We observed that mean/mode fill produces better results when compared to other data cleansing techniques such as combinational completion in terms of accuracy in

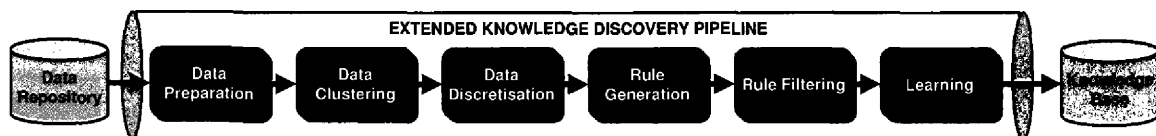


Figure 1: The extended knowledge discovery pipeline

view that the mean/mode fill substitutes missing values with the mean value for numerical attributes of all observed entries for that attribute. For string attributes, missing values are substituted by the mode value, i.e. the most frequently occurring value among the observed entries for that attribute. In contrast, combinational completion is perceived to be more complex in view that it expands each missing value for each object into the set of possible values. That is, an object is expanded into several objects covering all possible combinations of the object's missing values. This could lead to the number of possible combination for objects with multiple missing values to grow very rapidly.

2.2 Stage II: Data Clustering

With un-annotated datasets in mind, this next phase involves the utilization of a novel clustering ensemble algorithm to annotate the previously un-annotated dataset. This step involves the application of Kohonen's SOM as the basic technique for clustering the un-annotated dataset. It is not our aim to explore in detail any enhancements to SOMs. However, here, we aim to boost the results of the SOM clustering through re-sampling. Training instances that were wrongly predicted in the k th classifier will play a more important role in the $k+1$ th classifier to produce optimum clustering results, i.e. the best possible annotation of the previously un-annotated dataset. We argue that the application of boosting and SOM techniques in a clustering mechanism is novel in view that it is more commonly applied within the context of other neural networks applications. Moreover, previous work on clustering ensembles focused more on using aggregation (Dimitriadou et al., 2003) and hypergraph partitioning (Strehl and Ghosh, 2002) strategies.

2.3 Stage III: Data Discretization

Following the clustering stage, we then discretize the annotated results. The objective of discretization is basically to clearly differentiate between continuous values that are likely to be present in the dataset. Continuous values that lie within each interval are then mapped to the same discrete value. This process would result in better rule generation. Here, we use Boolean Reasoning and we have observed that it best suits our purpose over other techniques that we have explored such as the Entropy/Minimum Description Length (MDL) technique. This is because Boolean Reasoning computes a larger interval range compared to Entropy/MDL. The interval range computed by

Entropy/MDL is 0.1, e.g. 1.15-1.25. We believe that small interval range will not be sufficiently succinct and practical to discretize data as there will be very few datasets that are categorised in that interval. In this sense, Boolean Reasoning computes larger interval cuts which lead to a lower number of categories being computed and, hence, allow clearer rules to be generated in the next stage.

2.4 Stage IV: Rule Generation

For the data mining proper, we choose to carry out rule generation as the product of the pipeline based on rough set analysis. We compute reducts using genetic algorithm, a technique that is widely used for this purpose in rough set analysis. We then look into a number of sub-tables which will be randomly sampled from the datasets. Proper reducts are computed from each of these samples. The reducts that occur the most often across these sub-tables are in some sense the most stable and can be categorised as dynamic reducts. Based on the computation of the reducts, rules will be generated. This would constitute knowledge as it is potentially capable of capturing complex relationship between attributes and decision values of the dataset.

2.5 Stage V: Rule Filtering

This rule filtering phase follows the induction of rules. This stage basically cleanses the output of the rule generation stage to ensure that weak rules are removed. For this purpose, we define a quantitative rule evaluation technique leading towards the definition of what we call a Rule Quality Function which would indicate the quality of an induced rule. Here, the Rule Quality Function is based on the support, consistency and coverage measurements used by Michalski (1983). After determining the quality of the rules, the actual filtering process can then be carried out. We observed that statistical methods like the mean and Receiver Operating Characteristic curve (ROC) serves as a good indicator in the rule filtering. This can be viewed as an initial step towards ensuring that the rule base generated by the rule induction process using the rough set theory is of good quality.

2.6 Stage 6: Learning

The link between this learning phase and the rule filtering phase is novel where the filtered rules will be trained. We argue that the inclusion of this learning stage to the overall knowledge discovery pipeline is novel in view that most pipelines would

stop after rule filtering. Here, we employ NNE techniques to effectively learn the filtered rules. This is done by using different neural network algorithms, i.e. Multilayer Perceptron, Generalized Feedforward Network, Modular Neural Network and Radial Basis Function Networks. For this NNE, the bagging technique would be employed and this involves re-sampling or re-weighting the training instances where all the results of each neural network are aggregated to produce better predictions. Here, we have done the aggregation via averaging. This will produce a neural knowledge base (weights) that would be considered the best abstraction of the knowledge from the rules and would serve as a robust repository for decision support even in the event that users do not provide sufficient input.

3 CONCLUSION

We would like to highlight that the knowledge discovery pipeline leaves much to be explored in terms of the algorithms and techniques that can be applied at each stage of the pipeline. For our extended pipeline, we have proposed mean/mode fill for data preparation, clustering ensemble with SOM for clustering or data annotation, Boolean Reasoning for discretization, rough set analysis for rule extraction, Rule Quality Function (based on support, consistency and coverage) for rule filtering and finally, a NNE for rule learning. In addition to contributing towards an extended knowledge discovery pipeline, we believe that this featured work will provide an alternative inductive approach to support diagnosis and decision support especially in the medical domain.

With the incorporation of ensembling techniques for clustering and learning, we hope to minimize the need, or to reduce the temptation, to switch to other algorithms by making the most out of the selected algorithm, i.e. SOM and our 'cocktail' of neural network algorithms. We are currently evaluating each stage of our extended knowledge discovery pipeline using continuous, discrete and also possibly mixed (continuous and discrete) medical datasets such as those on breast cancer and thyroid disease. We will also be exploring other methods of clustering ensembles in the second stage to be integrated into our extended pipeline in future. We believe this extended pipeline would result in more accurate knowledge-based predictions in our effort to make medical diagnosis and decision support more reliable and trustworthy.

REFERENCES

- Abidi, S.S.R. and Hoe, K.M., 2002. Symbolic Exposition of Medical Data-Sets: A Data Mining Workbench to Inductively Derive Data-Defining Symbolic Rules. In *Proceedings of the 15th IEEE Symposium on Computer Based Medical Systems (CBMS 2002)*. Maribor, Slovenia.
- Breiman, L., 1996. Bagging Predictors. *Machine Learning*. Vol. 24, pp. 123-140.
- Dimitriadou, E., Weingessel, A., and Hornik, K., 2003. A cluster ensembles framework. In Abraham, A., Köppen, M., and Franke, K. (eds.), *Design and Application of Hybrid Intelligent Systems*. Frontiers in Artificial Intelligence and Applications. Vol. 104, pp. 528-534. IOS Press.
- Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P., 1996. From Data Mining to Knowledge Discovery in Databases. *AI Magazine*. Vol. 17, No. 3, pp. 37-54.
- Freund, Y. and Schapire, R.E., 1995. A Decision-Theoretic Generalization of On-line Learning and an Application to Boosting. In *Proceedings of the Second European Conference on Computational Learning Theory*. Barcelona, Spain, pp. 23-37.
- Hansen, L.K. and Salamon, P., 1990. Neural Network Ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 12, pp. 993-1001.
- Johnson, D.S., 1974. Approximation Algorithms for Combinational Problems. *Journal of Computer and System Sciences*, Vol. 9, pp. 256-278.
- Michalski, R.S., 1983. A Theory and Methodology of Inductive Learning. In Michalski, R., Carbonell, J. and Mitchell, T. (eds.), *Machine Learning. An Artificial Intelligence Approach*, pp. 83-134. Springer-Verlag.
- Risvik, K.M., 1997. Discretization of Numerical Attributes - Preprocessing for Machine Learning. *Project Report*. Knowledge Systems Group, Department of Computer Systems and Telematics, Norwegian Institute of Technology, University of Trondheim, Norway.
- Strehl, A., and Ghosh, J., 2002. Cluster ensembles - a knowledge reuse framework for combining multiple partitions. *Journal on Machine Learning Research*. Vol. 3, pp. 583-617.
- Yang, Y and Kamel, M., 2003. Clustering Ensemble Using Swarm Intelligence. In *IEEE Swarm Intelligence Symposium*. Indianapolis, Indiana, USA.
- Zhou, Z.H., Jiang, Y. and Chen, S.-F., 2003. Extracting Symbolic Rules from Trained Neural Network Ensembles. *AI Communications*. Vol. 16, No. 1, pp. 3-15.



Full Paper Submission: 12th November 2004

Position Paper Submission: 26th November 2004

Author Notification: 10th January 2005

Final Camera-Ready Submission and Registration: 31st January 2005

7th International Conference on Enterprise Information Systems

Call for Papers

ICEIS 2005 will be organized by INSTICC, in collaboration with Florida International University - Miami (USA).

SCOPE

The purpose of the 7th International Conference on Enterprise Information Systems (ICEIS) is to bring together researchers, engineers and practitioners interested in the advances and business applications of information systems. Five simultaneous tracks will be held, covering different aspects of Enterprise Information Systems Applications, including Enterprise Database Technology, Systems Integration, Artificial Intelligence, Decision Support Systems, Information Systems Analysis and Specification, Internet Computing, Electronic Commerce and Human Factors.

ICEIS focuses on real world applications; therefore authors should highlight the benefits of Information Technology for industry and services. Ideas on how to solve business problems, using IT, will arise from the conference. Papers describing advanced prototypes, systems, tools and techniques and general survey papers indicating future directions are also encouraged. Papers describing original work are invited in any of the areas listed below. Accepted papers, presented at the conference by one of the authors, will be published in the Proceedings of ICEIS. Acceptance will be based on quality, relevance and originality. Both full research reports and work-in-progress reports are welcome. There will be both oral and poster sessions.

Special sessions, dedicated to case-studies and commercial presentations, as well as tutorials dedicated to technical/scientific topics are also envisaged: companies interested in presenting their products/methodologies or researchers interested in holding a tutorial are invited to contact the conference secretariat. Information concerning the previous ICEIS editions (ICEIS'99, ICEIS 2000, ICEIS 2001, ICEIS 2002, ICEIS 2003, ICEIS 2004) can be found at <http://www.iceis.org>.

TOPIC AREAS / CONFERENCE TRACKS

Each of these topic areas is expanded below but the sub-topics list is not exhaustive. Papers may address one or more of the listed sub-topics, although authors should not feel limited by them. Unlisted but related sub-topics are also acceptable, provided they fit in one of the following main topic areas:

- 1. Databases and Information Systems Integration**
- 2. Artificial Intelligence and Decision Support Systems**
- 3. Information Systems Analysis and Specification**
- 4. Software Agents and Internet Computing**
- 5. Human-Computer Interaction**

AREA 1: Databases and Information Systems Integration

- Coupling and Integrating Heterogeneous Data Sources
- Enterprise Resource Planning
- Middleware Integration
- Legacy Systems
- Organisational Issues on Systems Integration
- Distributed Database Applications
- Object-Oriented Database Systems
- Enterprise-Wide Client-Server Architecture
- Database Security and Transaction Support
- Data Warehouses
- Multimedia Database Applications
- Web Databases
- Mobile Databases
- Software Engineering
- Software Measurement

AREA 2: Artificial Intelligence and Decision Support Systems

- Intelligent Agents
- Industrial Applications of Artificial Intelligence
- Strategic Decision Support Systems
- Group Decision Support Systems
- Applications of Expert Systems
- Advanced Applications of Fuzzy Logic
- Advanced Applications of Neural Network
- Natural Language Interfaces to Intelligent Systems
- Bayesian Networks
- Evolutionary Programming
- Coordination in Multi-Agent Systems
- Intelligent Social Agents and Distributed Artificial Intelligence Applications
- Intelligent Tutoring Systems
- Datamining
- Case-Based Reasoning Systems
- Knowledge-based Systems Engineering
- Knowledge Management

AREA 3: Information Systems Analysis and Specification

- Systems Engineering Methodologies
- Information Engineering Methodologies
- Organisational Semiotics
- Semiotics in Computing
- Requirements Analysis
- Ontology Engineering
- Modelling Formalisms, Languages, and Notations (e.g. UML, ER variants)
- CASE Tools for System Development
- Modelling of Distributed Systems
- Modelling Concepts and Information Integration Tools
- Business Processes Re-engineering
- Security, Freedom and Privacy

AREA 4: Software Agents and Internet Computing

- B2B and B2C Applications
- Process Design and Organisational Issues in e-Commerce
- E-Procurement and Web-based supply chain management
- Market-spaces: market portals, hubs, auctions
- E-Learning and e-Teaching
- Intranet and Extranet Business Applications
- Agents for Internet Computing
- Web Information Agents
- Case studies on Electronic Commerce
- Public sector applications of e-Commerce
- Interactive and Multimedia Web Applications
- Network Implementation Choices
- Object Orientation in Internet and Distributed Computing
- Internet and Collaborative Computing
- Semantic Web Technologies
- Wireless and Mobile Computing
- Agent-Oriented Programming

AREA 5: Human-Computer Interaction

- HCI on Enterprise Information Systems
- Functional and non-functional Requirements
- Internet HCI: Web Interfaces and Usability
- Design Methodology and Cognitive Factors in Design
- Multimedia Systems
- Machine perception: vision, speech, other
- Virtual Reality and Augmented Reality
- Intelligent User Interfaces
- User Needs
- Human Factors
- Accessibility to Disabled Users

- Geographical Information Systems
- E-Learning
- Computer Art

WORKSHOPS

- Wireless Information Systems
- Verification and Validation of Enterprise Information Systems
- Natural Language Understanding and Cognitive Science
- Ubiquitous Computing
- Security In Information Systems
- Computer Supported Activity Coordination
- Web Services: Modeling, Architecture and Infrastructure
- Organizational Modeling & Simulation Using Petri Nets
- Pattern Recognition in Information Systems

KEYNOTE SPEAKERS

- Richard Soley, Object Management Group, Inc, USA
- Jan Dietz, Delft University of Technology, The Netherlands
- Thomas Greene, MIT, USA
- Rosalind W. Picard, MIT, USA
- Henri Barki, HEC Montréal, Canada
- Daniel Schwabe, PUC-Rio, Brazil
- Lynne Markus, Bentley College, USA
- Raghavan Srinivas, Sun Microsystems, USA

TUTORIALS

- Eduardo B. Fernandez, Florida Atlantic University, USA

More to be available.

SUBMISSION OF PAPERS

Authors should submit a paper in English up to 5.000 words, using the paper format indicated below. The program committee will review all papers and the contact author of each paper will be notified of the result, by email. Each paper should clearly indicate the nature of its technical/scientific contribution, and the problems, domains or environments to which it is applicable. Authors must also indicate the conference track to which the paper is submitted. The paper must be carefully checked for correct grammar and spelling.

Authors should use the ICEIS web-based submission facility available at the conference web site, to enter the contact information of all paper authors plus the file containing the paper. The facility will automatically send a submission acknowledgement, by email, to the author indicated as "contact author".

A "blind" paper evaluation method will be used. To facilitate that, the authors are kindly requested to produce and provide the full paper, WITHOUT any reference to the authors. The manuscript must contain, in its first page, the paper title, an abstract and a list of keywords but NO NAMES OR CONTACT DETAILS WHATSOEVER are to be included in any part of this file.

If the author is unable to use the web-based procedure then he/she can send the paper by email to the **secretariat** attaching a second file containing: the title, author(s), affiliation(s), contact details, a list of keywords and an abstract.

The camera-ready format should be strictly used for all submitted papers. Please check and carefully follow the **instructions and templates provided**.

Due to space limitations in the Proceedings, the camera-ready version will be limited to 8 (eight) pages for full papers, 6 (six) for short papers (progress reports) and 4 (four) for poster presentations. If absolutely needed, it will be allowed to increase the total number of pages, by 4 extra pages. However, for each excess page the author will have to pay an additional fee.

IMPORTANT DEADLINES

Full Paper Submission: 12th November 2004

Position Papers Submission: 26th November 2004

Author Notification: 10th January 2005

Workshops Submission: 25th January 2005

Final Camera-Ready Submission and Registration: 31st January 2005

SECRETARIAT

ICEIS 2005 Secretariat

EC 3100

Department of Industrial & Systems Engineering

Florida International University

Miami, FL 33199

USA

E-mail for paper submission, proceedings and registrations: **secretariat@iceis.org**

E-mail for local arrangements: will be available soon

Web: **<http://www.iceis.org/>**

VENUE

Will be available soon.

CONFERENCE CO-CHAIRS

Chin-Sheng Chen, Florida International University, USA

Joaquim Filipe, Escola Superior de Tecnologia de Setúbal, Portugal (jfilipe@iceis.org)

PROGRAM CO-CHAIRS

Isabel Seruca, Universidade Portucalense, Portugal (iseruca@upt.pt)

José Cordeiro, Escola Superior de Tecnologia de Setúbal, Portugal (jcordeiro@iceis.org)

GENERAL WORKSHOPS CHAIRS

Slimane Hammoudi, École Supérieure d' Electronique de l' Ouest, France

(slimane.hammoudi@eseo.fr)

Olivier Camp, École Supérieure d' Electronique de l' Ouest, France (olivier.camp@eseo.fr)

SENIOR PROGRAM COMMITTEE

Luís Amaral, University of Minho, Portugal

Peter B. Andersen, Univ. of Aarhus, Denmark

Ricardo Baeza-Yates, University of Chile, Chile

Jean Bézivin, University of Nantes, France

Enrique Bonsón, University of Huelva, Spain

João Alvaro Carvalho, University of Minho, Portugal

Albert Cheng, Univ. of Houston and Rice Univ., USA

Miguel Delgado, University of Granada, Spain

Jan Dietz, Delft Univ. Technology, The Netherlands

Frank Dignum, Utrecht University, The Netherlands

António Figueiredo, University of Coimbra, Portugal

Göran Goldkuhl, Linköping University, Sweden

Thomas Greene, MIT, USA

Nuno Guimarães, University of Lisbon, Portugal

Jeet Gupta, University of Alabama in Huntsville, USA

Jean-Paul Haton, LORIA/INRIA, France

Alberto Laender, Federal Univ. of Minas Gerais, Brazil

Miguel Galejo, Declarativa, Portugal
Coral Calero, University of Castilla-La Mancha, Spain
Luis M. Camarinha-Matos, New University of Lisbon / Uninova, Portugal
Jorge Cardoso, University of Madeira, Portugal
Fernando Carvalho, Universidade Federal do Ceará, Brazil
Jose Jesus Castro-Schez, Universidad de Castilla-La Mancha, Spain
Luca Cernuzzi, Universidad Católica, Paraguay
Maria Filomena Cerqueira de Castro Lopes, Universidade Portucalense, Portugal
Elizabeth Chang, School of Information Systems, Australia
Laurent Chapelier, Banque Générale du Luxembourg, France
William Cheng-Chung Chu, TungHai University, Taiwan
Rodney Clarke, Staffordshire University, United Kingdom
Chrisment Claude, IRT/SIG, France
Francesco Colace, Università degli Studi di Salerno, Italy
Bernard Coulette, University of Toulouse 2, France
Sharon Cox, University of Central England, United Kingdom
Mohamed Dahchour, Institut National des Postes et Télécommunications (INPT), Morocco
Sergio de Cesare, Brunel University, United Kingdom
Andrea De Lucia, Università di Salerno, Italy
ino de Magalhães Ribeiro, Centro de Estudos e Recursos Multimediáticos (CEREM), UFP, Portugal
José Javier Dolado, University of the Basque Country, Spain
Jean-Christophe Dubois, Université de Rennes I, France
Schahram Dustdar, Technical University of Vienna, Austria
Alan Eardley, Staffordshire University, United Kingdom
David Emery, Staffordshire University, United Kingdom
Jean-Max Estay, I.M.A., France
João Faria, FEUP, Portugal
Jesus Favela, CICESE, United States of America
Eduardo Fernández-Medina, University of Castilla-La Mancha, Spain
Edilson Ferneda, Universidade Católica de Brasília, Brazil
Paulo Ferreira, INESC-ID/IST - Technical University of Lisbon, Portugal
Filomena Ferrucci, University of Salerno, Italy
Andrew Finegan, Charles Darwin University, Australia
Andre Flory, INSA de LYON, France
Donal Flynn, UMIST, United Kingdom
Ulrich Frank, Koblenz University, Germany
Ana Fred, IST - Technical University of Lisbon, Portugal
Lixin Fu, University of Carolina, United States of America
Juan Garbajosa, Technical University of Madrid, Spain
Joseph Giampapa, Carnegie Mellon University, United States of America
Raúl Giráldez, University of Seville, Spain
Pascual González, Universidad de Castilla-La Mancha, Spain
Robert Goodwin, Flinders University of South Australia, Australia
Silvia Gordillo, LIFIA-UNLP, Argentina
Feliz Gouveia, University Fernando Pessoa / CEREM, Portugal
Virginie Govaere, INRS, France
Jan Gulliksen, Uppsala University, Sweden
Rune Gustavsson, Blekinge Institute of Technology, Sweden
Sissel Guttormsen Schär, Swiss Federal Institute of Technology, Switzerland
Lamia Hadrich Belguith, FESGS, University of Sfax, Tunisia
Thorsten Hampel, University of Paderborn, Germany
Michael Heng, University of South Australia, Australia
Francisco Herrera, University of Granada, Spain
Colin Higgins, University of Nottingham, United Kingdom
Peter Higgins, Swinburne University of Technology, Australia
Erik Hollnagel, University of Linköping, Sweden
Jun Hong, Queen's University Belfast, United Kingdom
Nguyen Hong Quang, Institut de la Francophonie pour l'Informatique (IFI), Viet Nam
Jiankun Hu, RMIT University, Australia
Kaiyin Huang, K H Group BV, The Netherlands
Patrick C. K. Hung, University of Ontario Institute of Technology (UOIT), Canada

Maurizio Lenzerini, Univ. La Sapienza of Rome, Italy
Michel Léonard, University of Geneve, Switzerland
Kecheng Liu, The University of Reading, UK
Peri Loucopoulos, University of Manchester, UK
Paul Luker, Bournemouth University, UK
Kalle Lyytinen, Case Western Reserve University, USA
Yannis Manolopoulos, Aristotle U. Thessaloniki, Greece
José Legatheaux Martins, New Univ. Lisbon, Portugal
Masao Matsumoto, University of Tsukuba, Japan
James Odell, James Odell Associates , USA
George Papadopoulos, University of Cyprus, Cyprus
Luis Moniz Pereira, New Univ. Lisbon, Portugal
Alain Pirotte, Catholique Univ. of Louvain, Belgium
Klaus Pohl, University of Duisburg-Essen, Germany
Matthias Rauterberg, Technical University Eindhoven, The Netherlands
Colette Rolland, University of PARIS-1, France
Abdel-Badeeh Salem, Ain Shams University, Egypt
Bernadette Sharp, Staffordshire University , UK
Alexander Smirnov, St.Petersburg - SPIIRAS, Russia
Ronald Stamper, Univ. of Twente, The Netherlands
Reza Torkzadeh, University of Nevada Las Vegas, USA
Miguel Toro, University of Sevilla, Spain
José Tribolet, INESC/IST, Portugal
François Vernadat, European Commission, Luxembourg
Merrill Warkentin, Mississippi State University, USA
Hans Weigand, Tilburg University, The Netherlands
Roel Wieringa, University of Twente, The Netherlands

(list not yet complete)

REGULAR PROGRAM COMMITTEE

Jesus S. Aguilar-Ruiz, University of Seville, Spain
Patrick Albers, ESEO - Ecole Supérieure d'Electronique de l'Ouest, France
Salah Al-Sharhan, Gulf University for Science and Technology, Kuwait
Andreas Andreou, University of Cyprus, Cyprus
Pedro Antunes, Faculty of Sciences, University of Lisboa , Portugal
Joaquim Aparicio, New University of Lisbon, Portugal
Juan Carlos Augusto, (University of Ulster at Jordanstown, UK), United Kingdom
Bart Baesens, School of Management, University of Southampton, United Kingdom
Cecilia Baranauskas, UNICAMP - Universidade Estadual de Campinas, Brazil
Balbir Barn, Thames Valley University, United Kingdom
Senén Barro, University of Santiago de Compostela, Spain
Remi Bastide, LIHS - IRIT, France
Nadia Bellalem, University NANCY 2, France
Peter Bernus, Griffith University, Australia
Peter Bertok, RMIT University, Australia
Robert Biddle, Carleton University, Canada
Oliver Bittel, FH Konstanz - University of Applied Sciences, Germany
Fernando Boavida, University of Coimbra, Portugal
Luis Borges Gouveia, University Fernando Pessoa, Portugal
Djamel Bouchaffra, Oakland University, Michigan, United States of America
Danielle Boulanger, University of Lyon, France
Jean-Louis Boulanger, University of Technologie of Compiègne, France
José Ângelo Braga de Vasconcelos, Universidade Fernando Pessoa, Portugal

Hamid Jahankhani, University of East London, United Kingdom
Arturo Jaime, Universidad del Pais Vasco, Spain
Luis Jiménez Linares, UCLM, Spain
Luis Joyanes, Universidad Pontificia de Salamanca, Spain
Nikos Karacapilidis, University of Patras, Greece
Dimitris Karagiannis, University of Vienna, Austria
Stamatis Karnouskos, Fraunhofer, Germany
Hiroyuki Kawano, Nanzan University, Japan
Nicolas Kemper Valverde, Universidad Nacional Autónoma de México, Mexico
Manuel Kolp, University of Louvain, Belgium
John Krogstie, SINTEF and IDI, NTNU, Norway
Stan Kurkovsky, Columbus State University, United States of America
Yannick Lallement, Novator Systems, Canada
Chul-Hwan Lee, University of Pittsburgh, United States of America
Carlos León de Mora, University of Sevilla, Spain
Hareton Leung, Hong Kong Polytechnic University, China
Therese Libourel, LIRMM, France
John Lim, National University of Singapore, Singapore
Jiirni Linna, University of Vaasa, Finland
Anders Ljungberg, Gothenburg University, Sweden
Stephane Loiseau, LERIA, France
João Correia Lopes, University of Porto, Portugal
María Dolores Lozano, University of Castilla-La Mancha, Spain
Jianguo Lu, University of Windsor, Canada
Christopher Lueg, Charles Darwin University, Australia
Edmundo Madeira, UNICAMP - University of Campinas, Brazil
Laurent Magnin, University of Montreal, Canada
Mirko Malekovic, FOI - Zagreb University, Croatia
Nuno Mamede, L2F INESC-ID Lisboa / IST, Portugal
João Bosco Mangueira Sobral, UFSC / CPGCC - Computer Science Pos-Graduate, Brazil
Esperanza Marcos, Universidad Rey Juan Carlos, Spain
Farhi Marir, London Metropolitan University, United Kingdom
Maria João Marques Martins, IST - Technical University of Lisbon, Portugal
Herve Martin, LSR-IMAG, France
Johannes Mayer, University of Ulm, Germany
Andreas Meier, University of Fribourg, Switzerland
Emilia Mendes, The University of Auckland, New Zealand
Engelbert Mephu Nguifo, Université d'Artois - IUT de Lens, CRIL-CNRS, France
Miguel Mira da Silva, Instituto Superior Técnico, Portugal
Jouko Mäkelä, University of Vaasa, Finland
Paula Morais, Universidade Portucalense, Portugal
Fernando Moreira, Universidade Portucalense, Portugal
José Moreira, Universidade Portucalense, Portugal
Hector Munoz-Avila, Lehigh University, United States of America
Ana Neves, knowman - Consultadoria em Gestao, Lda, Portugal
Jose Angel Olivas, University of Castilla-La Mancha, Spain
Luis Olsina Santos, Universidad Nacional de La Pampa, Argentina
Peter Oriogun, London Metropolitan University, United Kingdom
Marcin Paprzycki, Oklahoma State University, United States of America
José R. Paramá, University of A Coruña, Spain
Oscar Pastor, DSIC-UPV, Spain
Maria Carmen Penadés Gramaje, Technical University of Valencia, Spain
Gabriel Pereira Lopes, Universidade Nova de Lisboa, Portugal
Laurent Péridy, IMA-UCO, France
Antonio Pescapé, University of Napoli, Italy
Stefan Peters, Vrije Universiteit Amsterdam, The Netherlands
Paolo Petta, Austrian Research Institute for Artificial Intelligence, Vienna, Austria
José Adriano Pires, Escola Superior de Tecnologia e Gestão / IPB, Portugal
Jacek Plodzien, Polish Academy of Sciences, Poland
Geert Poels, Faculty of Economics and Business Economics, Ghent University, Belgium
Macario Polo, University of Castilla-La Mancha, Spain
Bhanu Prasad, Florida A & M University, United States of America

Ed Price, Georgia Tech, United States of America
Pedro Ramos, ISCTE, Portugal
Ulrich Reimer, Business Operation Systems, Switzerland
Marinette Revenu, ENSICAEN, France
Simon Richir, CPNI Lab., France
António Rito-Silva, INESC/IST-Technical University of Lisbon, Portugal
David Rivreau, Université Catholique de l'Ouest, France
Pilar Rodriguez, Universidade Autónoma de Madrid, Spain
Agostinho Rosa, IST - Technical University of Lisbon, Portugal
Gustavo Rossi, LIFIA-UNLP, Argentina
Narcyz Roztockí, State University of New York (SUNY) at New Paltz, United States of America
Francisco Ruiz, Universidad de Castilla-La Mancha, Spain
Henry Samier, Université d'ANGERS, France
Manuel Santos, University of Minho, Portugal
Daniel Schang, ESEO, France
Arno Scharl, University of Western Australia, Australia
Mareike Schoop, Aachen University of Technology , Germany
Hanifa Shah, Staffordshire University, United Kingdom
Jianhua Shao, Cardiff University, United Kingdom
Timothy K. Shih, Tamkang University , Taiwan
Charles Shoniregun, University of East London, United Kingdom
Alberto Silva, INESC/IST - Technical University of Lisbon, Portugal
Maria João Silva Costa Ferreira, Universidade Portucalense, Portugal
Janice Sipiør, Villanova University, United States of America
Hala Skaf-Molli, INRIA Lorraine - University Henri Poincaré, France
Chantal Soule-Dupuy, University of Toulouse 1 - IRT, France
Chris Stary, University of Linz, Austria
Vijayan Sugumaran, Oakland University, United States of America
Lily Sun, Staffordshire University, United Kingdom
David Taniar, Monash University, Australia
Sotirios Terzis, University of Strathclyde, United Kingdom
Philippe Thiran, Eindhoven University of Technology, The Netherlands
Claudine Toffolon, Université du Littoral - Côte d'Opale, France
Robert Tolksdorf, Freie Universität Berlin, Germany
Ambrosio Toval, University of Murcia, Spain
Guiden Uchyigit, Imperial College, United Kingdom
Antonio Vallecillo, Universidad de Málaga, Spain
Luminita VasIU, Westminster University, School of Computing Science, United Kingdom
Christine Verdier, INSA of Lyon, France
Maria-Amparo Vila, University of Granada, Spain
HO Tuong Vinh, Institut de la Francophonie pour l'Informatique (IFI), Viet Nam
Aurora Vizcaino, Escuela Superior de Informática, Spain
Frank Wang, London Metropolitan University, United Kingdom
Hans Weghorn, University of Cooperative Education, Stuttgart, Germany
Gerhard Weiss, Technische Universitaet Muenchen, Germany
Graham Winstanley, University of Brighton, United Kingdom
Wita Wojtkowski, Boise State University, United States of America
Robert Wrembel, Poznan University of Technology, Poland
Baowen Xu, Southeast University, China
Hongji Yang, De Montfort University, United Kingdom
Yoneo Yano, Tokushima University, Japan
Kokou Yetongnon, University of Bourgogne , France
Liping Zhao, The University of Manchester, United Kingdom
Shuigeng Zhou, Fudan University, China
Lin ZongKai, Chinese Academy of Sciences, China

(list not yet complete)