Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia CIRP 7 (2013) 193 – 198

www.elsevier.com/locate/procedia

Forty Sixth CIRP Conference on Manufacturing Systems 2013

Quality Prediction in Interlinked Manufacturing Processes based on Supervised & Unsupervised Machine Learning

Daniel Lieber[a]*, Marco Stolpe[b], Benedikt Konrad[a], Jochen Deuse[a], Katharina Morik[b]

[a]*Institute of Production Systems, TU Dortmund, Leonhard-Euler-Str. 5, 44227 Dortmund, Germany*
[b]*Chair of Artificial Intelligence, TU Dortmund, 44221 Dortmund, Germany*
* Corresponding author. Tel.: +49-231-755-2796; fax: +49-231-755-2649. *E-mail address*: daniel.lieber@tu-dortmund.de.

Abstract

In the context of a rolling mill case study, this paper presents a methodical framework based on data mining for predicting the physical quality of intermediate products in interlinked manufacturing processes. In the first part, implemented data preprocessing and feature extraction components of the Inline Quality Prediction System are introduced. The second part shows how the combination of supervised and unsupervised data mining methods can be applied to identify most striking operational patterns, promising quality-related features and production parameters. The results indicate how sustainable and energy-efficient interlinked manufacturing processes can be achieved by the application of data mining.

Keywords: Inline Quality Prediction; Data Preprocessing; Feature Subset Selection; Data Mining; Supervised Machine Learning; Sustainability; Ressource- and Energy-Efficiency

1. Introduction

Steel industry production is characterized by highly resource-consuming, complex and automated interlinked manufacturing processes. Technological and temporal restrictions limit physical product quality inspections to the final process step. Hence, undetected quality deviations passing through the entire value chain have severe impact on internal failure costs due to increasing rejection and reworking.

In this respect, recently *Alwood* and *Cullen* [1] presented remarkable research results summarizing that in 2008 60% (334 million tons) of world-wide re-melted steel scrap (574 mill. tons) never reached final products but was scrapped in advance. In contrast to 30% (98 mill. tons) accruing from steelmaking and casting processes, 70% (236 mill. tons) of production-related scrap can be associated with waste products resulting at later stages due to offcuts, surplus and defects.

New solutions for continuous quality monitoring are therefore needed and investigated in the context of a hot rolling mill case study, provided by a leading German steel company (see Fig. 1). In this context, nowadays

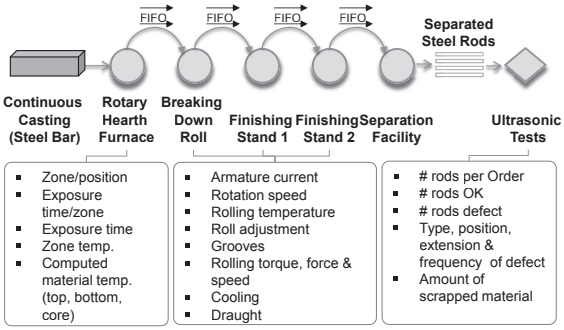


Fig. 1. Hot rolling mill process chain and recorded series data [5]

internal material defects cannot be monitored by external state-of-the-art sensors. Therefore, the goal is to identify quality deviations as early as possible and in real-time by data mining on distributed sensor measurements along the process chain. More specifically, since final product's quality depends on how it was processed, value series of sensor measurements recorded at each processing step might contain quality-related patterns. Given also quality labels generated from ultrasonic tests, supervised learning may derive prediction models which can predict the quality related physical properties of a

doi:10.1016/j.procir.2013.05.033

product already at intermediate production steps. Early detection of defects will save production resources and lead to more sustainable and energy-efficient interlinked manufacturing processes.

Previous publications have discussed the general challenges of distributed data mining on sensor data of interlinked processes [2] and the problem of deriving appropriate quality labels [3, 4]. Moreover, in [5] the implementation of a data acquisition and storage system was described and first prediction results on data recorded at the rotary hearth furnace were presented.

As now value series data from all processing units is available, there are three key contributions of this paper:

- First, an approach for automatically preprocessing value series data, extracting features and compiling them into a table format which is suitable for many supervised and unsupervised data mining algorithms is presented.
- Second, it is shown that by combining supervised and unsupervised data mining methods striking patterns in rolling mill process data can be identified, e.g. operational modes.
- Third, first results on predicting the final quality of steel bars and a selection of promising features for quality prediction are presented.

The remainder of this paper is arranged as follows: In the next section, related work regarding the state of the art of data mining in manufacturing is presented. Section 3 gives a general overview of the Inline Quality Prediction (IQP) System realized as automated processes in the open source data mining software RapidMiner [6]. Subsequently, section 4 introduces implemented system components focusing on data preprocessing and feature extraction. Followed by an overview of data mining tasks and algorithms in section 5, section 6 provides results of several analysis steps conducted on the preprocessed data. Finally, we conclude and discuss forthcoming work in section 7.

2. Data Mining in Manufacturing Processes

Literature reviews such as *Choudhary et al*. [7] illustrate the diversity of data mining techniques used in industrial manufacturing applications, like process characterization, reporting, fault diagnostics, product development, production scheduling as well as preventive maintenance, defect prediction and decision support systems. However, contrary to the increasing interest in manufacturing related knowledge discovery activities, data miner surveys by *Rexer Analytics* [8] or *KDnuggets* [9] annually show that production related data mining projects in manufacturing are still underrepresented in comparison to more well-established fields, e.g. consumer-analytics, fraud-detection or banking.

Focusing particularly on the steel industry, related work can be found by *Peters et al.* [10] who are conducting research on data mining techniques with focus on cause analyses, e.g. surface defects of flat steel products. In contrast to *Peters et al.*, the intention of the project presented in this paper is to develop universal distributed data mining techniques for real-time inline quality prediction focusing on detecting internal material defects that nowadays cannot be monitored by external state-of-the-art sensors.

3. Inline Quality Prediction (IQP) System

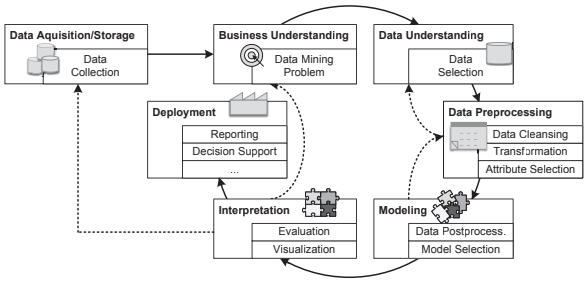


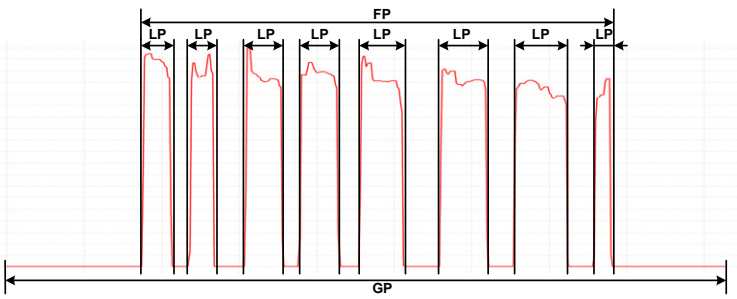
Fig. 2. Life cycle of a data mining project

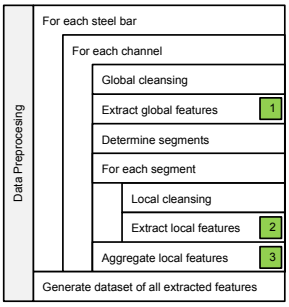
A final goal of the hot rolling mill case study is the design of a distributed Inline Quality Prediction (IQP) system that integrates and automates all necessary data mining steps, analyzing and classifying process patterns in real-time. Fig. 2 shows common data mining steps that are defined and summarized in the Cross Industry Standard Process for Data Mining (CRISP-DM) [11] and the widespread methodology of Knowledge Discovery in Databases (KDD) [12].

While the mining problem and the data acquisition component of the system were already discussed in previous work, the next section describes the general design of components for preprocessing and extracting features from value series, transforming them into a format that is suitable for many well-known data mining algorithms. The modular design is easily extensible and has been implemented as automated processes in RapidMiner. Future goals are already respected, like the distributed deployment of all preprocessing and feature extraction processes across real production units.

4. Data Preprocessing

Different types of sensor measurements like rolling force, speed and temperature (see Fig. 1) are transmitted in separate signal channels. For each steel bar, the value series of all recorded channels must be individually preprocessed. Fig. 3 shows the general steps for preprocessing on the upper right.





Resulting Feature Set

∀ bar_id n ∀ channel m		Channel 1												...	Channel m		
		Global Preprocessing (GP)				Local Preprocessing (LP)				Feature Postprocessing (FP)				...	GP / LP / FP		
Meta-Info		1 *Global Features (G)*				2 *Local Features (L)*				3 *Aggregated Features (A)*				...	1	2	3
bar_id	...	length	max	mean	...	length_seg	max_seg	mean_seg	...	differenz_mean_seg(i)_seg(i+1)	stddev_differenz_mean	mean_diff_mean_process	...	...	...	...	...
12345	...	330	0.410	0.161	...	44	0.379	0.356	...	-0.022	0.091	0.011	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
54321	...	342	0.425	0.163	...	68	0.394	0.366	...	-0.009	0.089	0.013	...	...	...	...	...

Total number of features p

Fig. 3. Resulting data table including meta data and features extracted from value series data of channels 1 to m

The steps are realized as a master process that executes different types of sub-processes (global and local cleansing, global and local feature extractions, etc.) for each bar and signal channel. These automatically call more generic procedures for preprocessing, like filtering, outlier cleansing or normalization (Table 1). These procedures can be reused by different processes of the system, providing specific parameters like thresholds depending on the particular channel.

Table 1. Excerpt of available preprocessing procedures

Generic Preprocessing Procedures
Outlier smoothing based on thresholds
Data cleansing by filtering
Rounding
Normalization
Segmentation

Fig. 3 shows an example series consisting of rolling force measurements, recorded during the processing of a single steel bar. Global cleansing involves removing irrelevant values where the steel bar was not processed. Since the sensors measure continuously, such intervals would be hard to detect by an automated process without knowledge about the domain.

The cleansing therefore relies on the meaning of values in other channels. For instance, a roll position above a certain threshold indicates no processing, allowing to remove values from the same interval in other channels. Similarly, the segmentation of series into meaningful intervals, like individual rolling steps, is easy visually, but sometimes hard to automate. Although methods for the automatic segmentation exist [13], it is much more reliable and faster to count rolling steps with a sensor and use that information for the splitting.

Table 2. Excerpt of available feature extraction procedures

Feature Extraction Procedures
Time-related statistics (length of the series or of segments)
Value statistics (min./max. values, sum, mean, median, std.deviation)
Statistics on differences between following values (min., max., etc.)
Statistics on differences between start and end values (min., max., etc.)
Measures for the deviation between target and actual values
Frequency statistics and counts (histograms, number of segments)

Global features (G) are extracted from the whole series right after global cleansing, while local features (L) are extracted from individual segments after a local preprocessing step (LP). Table 2 shows an excerpt of the types of features that can be calculated from series values. In a postprocessing step (FP), extracted features from segments may be aggregated further by the same methods. At the end, features extracted from each channel for each steel bar are collected in a single data table (see the lower part of Fig. 3), together with quality labels ("OK" or "NOK") for each individual steel bar, as generated from ultrasonic test records [4]. The resulting table, in this case consisting of 470 rows and 2,170 extracted features, can be analyzed by different data mining algorithms introduced in the next section.

The reasoning behind extracting aggregated statistics like the median or standard deviation is that if there were quality-related issues already on a large scale, the costly extraction of features describing fine-grained differences can be avoided. Moreover, knowledge on global patterns may support the identification of local ones.

5. Modeling

In the following, an overview of basic data mining tasks and algorithms is given and the methods used in the experimental section are introduced.

5.1. Unsupervised Learning and Clustering

Unsupervised methods, like visualization, clustering, outlier detection or dimension reduction are often used as a first step in data mining for getting an impression of patterns and relationships in complex data sets.

Clustering algorithms divide a set of observations into groups (clusters) such that observations inside clusters are more similar to each other than to those in other clusters. Thus, clustering can provide a first insight into similarity relationships. The feature values that were extracted according to section 4 are all numeric and can be interpreted as points in a metric space. Given a set $X = \{\vec{x}_1, \dots, \vec{x}_n\}$ of n observations represented by data points in a p-dimensional Euclidean space, i.e. $\vec{x}_i = (x_{i1}, \dots, x_{ip})$, $\vec{x}_i \in \mathbb{R}^p$, the dissimilarity between two observations $\vec{x}_a$ and $\vec{x}_b$ can, for instance, be measured by the Euclidean distance.

The *k-Means* algorithm [14] computes a partitioning of X into a user-specified number of k clusters, with the goal of minimizing the average Euclidean distance between the points in each cluster. Initially, it assigns all data points randomly to k clusters. Then, in alternating steps, it calculates the mean vector of each cluster and (re)assigns each data point to the cluster whose mean is closest, until the means do not change anymore. The algorithm should be started several times, since it can only find a locally optimal clustering.

The *Self Organizing Map* (SOM) [15] achieves a clustering and a dimensionality reduction by mapping input vectors from a higher dimensional continuous space to a fixed number of points (neurons) on a lower dimensional grid. Since similar input vectors are also lying close to each other on the grid, similarity relationships between high-dimensional input vectors can easily be visualized on a two dimensional map.

5.2. Supervised Methods and Classification

Supervised methods are used if some observations can already be labeled according to a known target concept and a rule or function (also called model) for this assignment should be learned from the data. Given distinct class labels, the *Naïve Bayes* [16] classifier predicts the class of a given observation x_i based on the prior probability $p(C)$ of the class and $p(x_{i1}, \dots, x_{ip}|C)$, the likelihood of the feature values given the class C:

$$p(C|x_{i1}, \dots, x_{ip}) = \frac{1}{Z} p(C) \prod_{j=1}^{p} p(x_{ij}|C) \quad (1)$$

Z is a normalization constant and can be ignored for classification. The probabilities $p(C)$ and $p(x_{ij}|C)$ are estimated from the set of labeled observations. Predicted is usually the class with the highest estimated probability. Although the method assumes the often not given independence of features it has achieved sufficient prediction accuracy in many practical applications.

Decision trees [17] classify observations by sorting them into axis parallel rectangular regions of the input space. The method recursively determines features whose values can be used for sorting observations into regions that contain as many points of the same class as possible. The actual classification is then performed by tests on the chosen features and their values, along a path from the root to the leaves of the tree.

The *nearest neighbor* method (k-NN) [18] stores a set of labeled observations. New observations are classified by majority vote of the k nearest neighbors.

The *support vector machine* (SVM) [19] determines the support vectors of a hyperplane which separates the observations of two classes with maximum margin. Providing a kernel function that measures the similarity of observations in a higher dimensional feature space allows for a non-linear separation of observations in the original input space.

5.3. Feature Selection

Often only a subset of features is relevant. In the case of distance based methods, noisy or highly correlated features can even disturb learning. If the selection of features is not already *embedded* in an algorithm, *filter* or *wrapper* approaches may be used for automatically determining a subset of relevant features [20]. While filters select features based on simple criteria, the wrapper approach evaluates generated subsets of features by training and applying a prediction model. In Section 6, an evolutionary wrapper approach is used for the selection of subsets. Starting with a population of random feature subsets, the algorithm successively combines and mutates the fittest individuals over several generations and returns the best found subset.

5.4. Evaluation

One possible measure for the accuracy of a model is the percentage of correctly classified observations. Maximizing the accuracy on the training set will usually underestimate the true error, which is the expected error also on unseen observations. A better estimation is the accuracy on an independent hold-out set not shown to the learning algorithm. If only few labeled observations are available, a standard procedure for estimating the accuracy is the n-fold *cross-validation* [21]. The training set is divided randomly into n equally sized subsets and each subset is used once as a hold-out set. The total accuracy is then the average of the accuracies on these hold-out subsets.

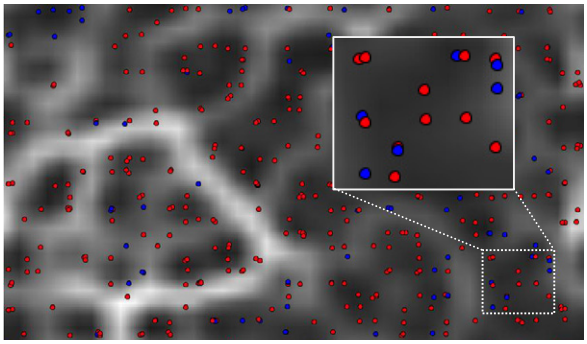


Fig. 4. Quality levels of steel bars plotted on a 40x30 SOM

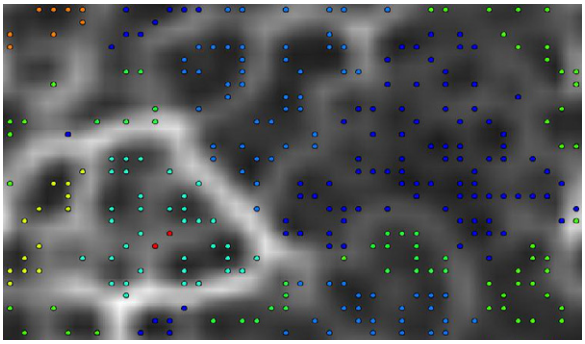


Fig. 5. Final product's end dimension plotted on a 40x30 SOM

6. Data Analysis and Results

For an analysis of the data set described in section 4, some of the extracted features are excluded. The block roll temperature is removed due to an unreliable sensor. Features on individual segments are not included since it is unclear how to compare different numbers of rolling steps in distance calculations. Therefore, only global and aggregated features are kept, resulting in 218 features describing each of the 470 processes.

6.1. Quality Prediction and Important Features

The feature vectors of the 470 processes (indicated by points) are mapped to different parts of a 40x30 SOM (see Fig. 4) for getting a first impression of the data. The color of a point represents the final quality (red for "OK" and blue for "NOK") of the resulting steel bar. Here, quality deviations are known to be porosities related to processing in the rolling mill. Processes that are similar according to their feature values are lying close to each other on the map. The distance between two points is additionally weighted as indicated by the shading, where lighter shades can be interpreted like hills on a geographical map. Dark areas surrounded by very light borders therefore indicate separate clusters of similar processes. In many cases, processes leading to a low final product quality are lying very close to those leading to a high quality (see e.g. the zoomed area in Fig. 4). A visual inspection of corresponding plots verifies that such series look almost identical on a large scale and that differences can rather be found in fine-grained details.

For assessing how well both classes can be separated quantitatively, the prediction accuracy of the supervised classification methods introduced in section 5 was estimated by a 10-fold cross-validation. Parameter settings and manually chosen subsets of features were varied in the experiments. Since k-NN looked most promising, an exhaustive evolutionary feature selection with k=7 over 50 generations was conducted and could achieve a prediction accuracy of 80.21%. The top 13 selected features are believed to be quality-relevant by domain experts and include the heating time in the hearth furnace, rolling force, speed and temperature. Because seven of the features regard statistics which measure differences within or between rolling steps, future work on improving prediction performance should focus on a more detailed description of these deviations.

Another promising result stems from focusing only on the correct prediction of high quality steel bars. Even though blue and red points are lying close together, the SOM in Fig. 4 also contains large continuous areas of processes resulting in high quality. If process parameters would stay in these ranges, avoiding those with higher probability for errors, it might already improve the process and reduce the amount of produced scrap metal.

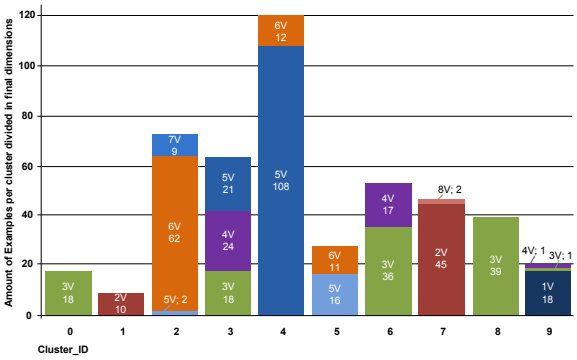


Fig. 6. Distribution of final product's end dimensions in 10 clusters

6.2. Detection of Distinct Operational Modes

Asking what processes look most similar to each other reveals a high similarity between processes resulting in the same final product's end dimension. Instead of plotting the quality, Fig. 5 shows the different end dimensions plotted in different colors on the SOM. Here, processes resulting in the same end dimension form large clusters, i.e. continuous separated areas of the same color. Further, Fig. 6 shows the distribution of end dimensions in each of ten clusters as determined by the k-Means algorithm. Many of the clusters include only

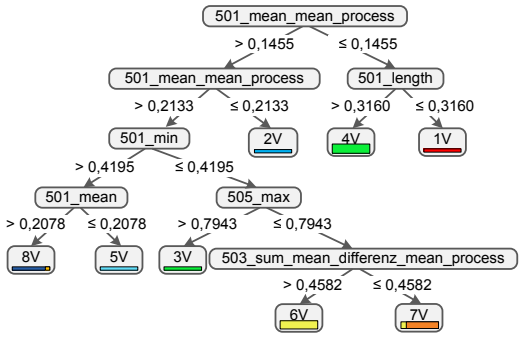


Fig. 7. Decision tree for detecting the final product's end dimensions

one or two end dimensions that are similar. A decision tree (see e.g. Fig. 7) trained solely on the features of the first finishing roll could decide on the end dimension with an accuracy of 90%, while k-NN (k=11) achieved about 97%, indicating a high correlation between end dimensions and processes. Most important for deciding on the product's final dimension is the position of the roll (channel 501) which makes sense, since it determines the height of the end product. Experts have verified that the mapping between large scale process behavior and the end dimension produced, detected by means of automated data mining methods, reflects the real modes of operation in the rolling mill.

7. Conclusion

Data mining on features of a rolling mill process is able to detect meaningful and striking operational patterns previously only known to domain experts. The result is an improvement on previous work [4] where, only based on static features from the rotary hearth furnace, no such patterns could be detected. The clustering of aggregated sensor measurements now allows for a quantitative description of deviations from the intended operational modes, enabling e.g. automated process monitoring. Moreover, first prediction results indicate only a low correlation between modes and the quality of the end product, verifying the correctness of operation on a large scale. The description of distinct modes could further give valuable hints for the detection of quality-related patterns that are independent from the final product's end dimension.

An exhaustive evolutionary feature selection indicates promising features for quality prediction, including heating times in the furnace and deviations of rolling force, speed and temperature within and between consecutive rolling steps. Future work will show if the extraction of more fine-grained deviations can lead to an improvement in prediction performance. Moreover, focusing only on the correct prediction of high-quality steel bars may help with learning a model for more safe and reliable process parameter ranges.

Acknowledgements. This work has been supported by the DFG, Collaborative Research Center 876, project B3, http://sfb876.tu-dortmund.de/.

References

[1] Alwood, J.M., Cullen, J.M., 2012. Sustainable Materials - With Both Eyes Open. UIT Cambridge, Cambridge, p. 51-54.
[2] Stolpe, M., Morik, K., Konrad, B., Lieber, D., Deuse, J., 2011. Next Generation Data Mining Summit 2011. Athens, Greece.
[3] Stolpe, M., Morik, K., 2011. Learning from Label Proportions by Optimizing Cluster Model Selection, in "ECML PKDD 2011", Gunopulos, D. and others, Part III, Vol. 6913, Springer, Berlin, Heidelberg, pp. 349-364.
[4] Konrad, B., Lieber, D., Deuse, J., 2012. Striving for Zero Defect Production: Intelligent Manufacturing Control through Data Mining in Continuous Rolling Mill Processes, in "Robust Manufacturing Control" K. Windt, Vol. 1, No. 1, Springer, Heidelberg, pp. tba.
[5] Lieber, D., Konrad, B., Deuse, J., Stolpe, M., Morik, K., 2012. Sustainable Manufacturing Processes through Real-Time Quality Prediction, in „Leveraging Technology for a Sustainable World", Dornfeld, D.A., Linke, B., Springer, Heidelberg, pp. 393-398.
[6] Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., Euler, T., 2006. YALE: Rapid Prototyping for Complex Data Mining Tasks, in „KDD´06: Proc. of the 12th ACM SIGKDD int. conf. on Knowledge Discovery and Data Mining", Ungar, L., Craven, M., Gunopulos, D., Eliassi-Rad, T., ACM, New York, pp. 935-940.
[7] Choudhary, A.K., Harding, J.A., Tiwari, M.K., 2009. Data Mining in Manufacturing: A Review Based on the Kind of Knowledge, Journal of Intelligent Manufacturing, Springer, Vol. 20, pp. 501-525.
[8] Rexer Analytics, http://www.rexeranalytics.com, 2012.
[9] KDnuggets, http://www.kdnuggets.com/polls, 2012.
[10] Peters, H., Ebel, A., Holzknecht, N., Link, N., Hackmann, J., Heckenthaler, T., Lücking, F., Pander, M., 2012. Industrial data mining in the steel industry, Stahl und Eisen, 132(2), pp. 29-37.
[11] Nisbet, R., Elder, J., Miner, G., 2009. Handbook of Statistical Analysis and Data Mining Applications, Elsevier, San Diego, p. 39.
[12] Fayyad, U., Piatetski-Shapiro, G., Smyth, P., 1996. From Data Mining to Knowledge Discovery in Databases, AI Magazine, Vol. 17, No. 3, pp. 37-54.
[13] Rakthanmanon, T., Keogh, E.J., Lonardi, S., Evans, S., 2012. MDL-based time series clustering, Knowl Inf Sys, Vol. 33, pp. 371-399.
[14] MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations, Symp. Math. Stat. & Prob., pp. 281-297.
[15] Kohonen, T., 2001. Self-Organizing Maps, Springer, 3rd ed.
[16] John, G.H., Langley, P., 1995. Estimating continuous distributions in Bayesian classifiers, in "Proc. of the 11th Conf. on Uncertainty in Artificial Intelligence", Morgan Kaufmann, San Francisco, pp. 338-345.
[17] Quinlan, J.R., 1986. Induction of decision trees, Machine Learning, Vol. 1, No. 1, pp. 81-106.
[18] Aha, D., 1992. Tolerating noisy, irrelevant, and novel attributes in instance-based learning algorithms, International Journal of Man-Machine Studies, Vol. 36, No. 2, pp. 267-287.
[20] Vapnik, V., 1999. The nature of statistical learning theory, Springer, New York, 2nd ed.
[21] Pudil, P., Novovicová, J., 1998. Novel Methods for feature Subset Selection with Respect to Problem Knowledge, in "Feature Extraction, Construction and Selection – A Data Mining Perspective" H. Liu, H. Motoda, Kluwer, Dortdrecht, pp. 101-116.
[22] Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, 2nd ed.