DISESOR - decision support system for mining industry

Michał Kozielski
Institute of Electronics,
Silesian University of Technology
ul. Akademicka 16,
44-100 Gliwice, Poland
Email: michal.kozielski@polsl.pl

Marek Sikora
Institute of Informatics,
Silesian University of Technology
ul. Akademicka 16,
44-100 Gliwice, Poland
Email: marek.sikora@polsl.pl

Łukasz Wróbel
Institute of Innovative Technologies EMAG
Leopolda 31,
40-189 Katowice, Poland
Institute of Informatics,
Silesian University of Technology
ul. Akademicka 16,
44-100 Gliwice, Poland
Email: lukasz.wrobel@ibemag.pl

Abstract—This paper presents the DISESOR integrated decision support system. The system integrates data from different monitoring and dispatching systems and contains such modules as data preparation and cleaning, analytical, prediction and expert system. Architecture of the system is presented in the paper and a special focus is put on the presentation of two issues: data integration and cleaning, and creation of prediction model. The work contains also a case study presenting an example of the system application.

I. INTRODUCTION

COAL mining is a heavy industry that plays an important role on an energy market and employs hundreds of thousands of people. Coal mining is also an industry, where large amount of data is produced but little is done to utilise them in further analysis. There is also a justified need to introduce a decision support system (DSS) integrating different aspects of coal mine operation in order to maintain continuity of mining.

Currently coal mines are well equipped with the monitoring, supervising and dispatching systems connected with machines, devices and transport facilities. There are also the systems for monitoring natural hazards (methane-, seismic- and fire hazards). All these systems are provided by many different companies, what causes problems with quality, integration and proper interpretation of the collected data. The collected data are used chiefly for current (temporary) visualisation on boards which display certain places in the mine. Whereas, application of domain knowledge and the results of historical data analysis can improve the operator's and supervisor's work significantly. For example, thanks to short-term prognoses about methane concentration, linked with the information about the location and work intensity of the cutter loader, it is possible to prevent emergency energy shutdowns and maintain continuity of mining (the research on this methodology was discussed in [1]). This will enable to increase the production volume and to reduce the wear of electrical elements whose exploitation time depends on the number of switch-ons and switch-offs.

It is possible to see the rising awareness of monitoring systems suppliers who begin to understand the necessity to make the next step in these systems development. Therefore, the companies providing monitoring systems seek their competitive advantage in equipping their systems with knowledge engineering, modelling and data analysis methods. This is a strong motivation to consider a DSS presented in this paper.

The goal of this paper is to present an architecture of the integrated decision support system DISESOR. The system integrates data from different monitoring systems and contains an expert system module, that can utilise domain expert knowledge, and analytical module, that can be applied to diagnosis of the processes and devices and to prediction of natural hazards. The special focus of the paper is put on the data integration and data cleaning issues realised by means of the data warehouse and ETL process. The work also contains a more detailed presentation of the prediction module, which is complemented by a presentation of a simple case of methane concentration prediction in a coal mine.

The contribution of the paper consists of:

- the architecture of the integrated decision support system DISESOR,
- presentation of the approaches to the preparation and cleaning of the data collected by monitoring systems,
- presentation of the prediction module architecture and principles of the module operation,
- case study presenting application of the presented system to methane concentration prediction in a coal mine.

The structure of the paper is as follows. Section II presents the works related to the presented topic. The architecture of the DISESOR system and its data repository are presented in section III. The more detailed descriptions of the data preparation and cleaning and prediction modules are presented in sections IV and V respectively. The case study of methane concentration prediction task is presented in section VI and section VII presents the final conclusions.

II. Related work

The typical environments deployed in a coal mine are monitoring and dispatching systems. These systems collect a large number of data which can be utilised in further analysis, e.g., on-line prediction of the sensor measurements, which area was surveyed in [2]. The analysis can address different aspects of coal mine operation such as, e.g., equipment failure or natural hazards.

The examples of the research in the field of natural hazards in an underground coal mine cover, e.g., methane concentration prediction and seismic hazard analysis. The research on the prediction of the methane concentrations was presented in [3, 1]. Application of data clustering techniques to seismic hazard assessment was presented in [4]. There are also approaches to prediction of seismic tremors by means of artificial neural networks [5] and rule-based systems [6]. Each research listed above is a stand alone approach not incorporated into any integrated system.

Analytical methods that were mentioned require the data which are extracted, cleaned, transformed and integrated. Decision support systems utilise a data repository of some kind, e.g., a data warehouse [7]. The critical dependence of the decision support system on a data warehouse implementation and an impact of the data quality on decision support is discussed in [8].

There are applications of machine learning methods to diagnostics of mining equipment and machinery presented in literature [9, 10]. Also some initial concepts of the system that processes data streams delivered by the monitoring systems were presented in [11]. However, to the best of the authors knowledge there is no example of the integrated decision support system for monitoring processes, devices and hazards in a coal mine (except the work dealing with DSS for coal transportation [12] which loosely corresponds to the given topic).

III. System architecture

The general architecture of the DISESOR integrated decision support system is presented in Fig. 1. The architecture of the system consists of data repository and data preparation and cleaning, that are presented in more detail in the following sections, and analytical, prediction and expert system modules shortly presented below, as they are not the main focus of the paper.

A. Decision support system

The core of analytical, prediction and expert system modules is based on the RapidMiner [13] platform. The RapidMiner environment was customised to the requirements of the non-advanced user by disabling unnecessary options and views. Therefore, an advanced user can use the whole functionality of RapidMiner, whereas the non-advanced user can use such thematic operators as e.g., "Solve a methane concentration prediction issue" or "Solve a seismic hazard issue". Also due to the target application of the system in Polish coal mines the RapidMiner environment was translated into Polish.

Finally, RapidMiner was extended in the created application by additional operators wrapping R [14] and MOA (Massive On-line Analysis) [15] environments.

The goal of the Data preparation and cleaning module, which is referred further as ETL2, is to integrate the data stored in data warehouse and process them to the form acceptable by the methods creating prediction and classification models. In other words the ETL2 module prepares the training sets.

Prediction module is aimed to perform incremental (on-line) learning of predictive models or apply classification and prediction models created in analytical module for a given time horizon and frequency of the values measured by the chosen sensors. This module also tracks the trends in the incoming measurements. The created predictive models are adapted to the analysed process on the basis of the incoming data stream and the models learnt on historical data (within the analytical module). The module provides the interfaces that enable the choice of quality indices and their thresholds that ensure the minimal prediction quality. If the quality of predictions meets the conditions set by a user, the predictions will be treated as the values provided by a soft sensor. They can be further utilised by e.g., expert system but also they can be presented to a dispatcher of a monitoring system.

Analytical module is aimed to perform analysis of historical data (off-line) and to report the identified significant dependencies and trends. The results generated by this module are stored in the repository only when accepted by a user. Therefore, this module supports a user in decision-making of what is interesting from monitoring and prediction point of view. It also provides additional information that can be utilised to enrich the knowledge of expert system or that can be utilised to comparative analysis. The module supports identification of changes and trends in the monitored processes and tools and it also enables to compare the operator's and dispatcher's work.

Expert system module is aimed to perform on-line and off-line diagnosis of machines and other technical equipment. It is also aimed to supervise the processes and to support the dispatcher or expert decision-making with respect to both technical condition of the equipment and improper execution of the process. The inference process is performed by means of classical inference based on stringent rules and facts or probabilistic inference based on belief networks. The system contains also a knowledge base editor that allows a user to define such rules and network. The expert system module is currently being developed.

B. Data repository

Data repository was designed as a data warehouse of a snowflake structure (as some dimensions have multiple levels), that is presented in Fig. 2 in a reduced, general form. The structure of a data warehouse results from the analysis of databases of the existing monitoring systems and the characteristics of the known sensors. The full list of tables with their description is presented in Table I.
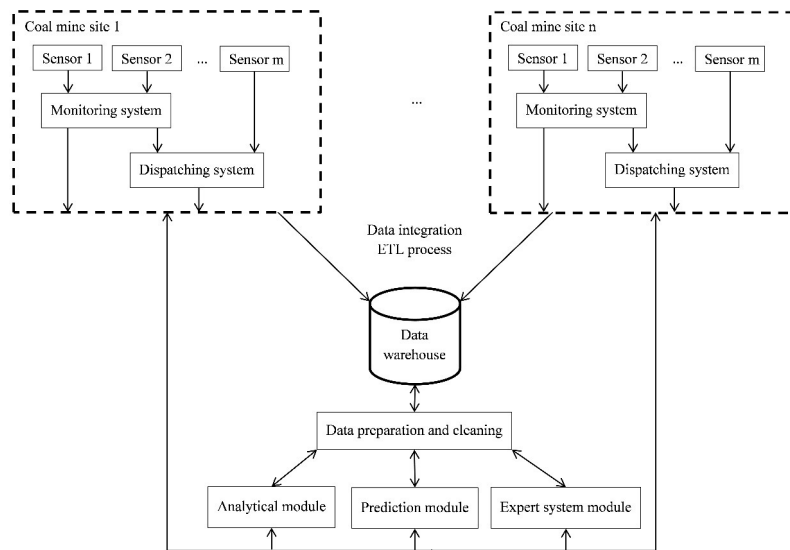
Fig. 1. Architecture of the DISESOR integrated decision support system

Table	Description
Measurement	Value of a measurement
State	State of a measurement, e.g., *alarm*, *calibration*, *breakdown*
Discretisation	The measured values can be of discrete type
Time	Time of a measurement, range $[00:00:00, 23:59:59]$, 1 second resolution
Time_category	Category, e.g., *mining* or *no mining*
Date	Date of a measurement
Location	Location of the measurement source
Location_attribute	Characteristics of the given location
Location_hierarchy	Hierarchical structure of location
Source	Measurement source, e.g., sensor or device
Source_attribute	Characteristics of the given source

TABLE I
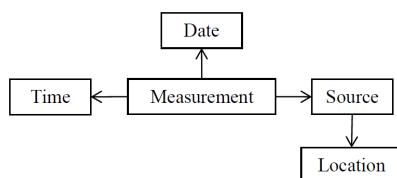TABLES CREATING A DATA WAREHOUSE STRUCTURE.

Fig. 2. Simplified schema of data repository

The central table of the data repository is *Measurement* where all the measurements are stored. The dimensions related to the *Measurement* table are *Date*, *Time* and *Source*. *Date* and *Time* describe when the measurement was registered, whereas *Source* describes what registered the given measurement. The *Source* table contains among others such information about sensors/devices as:

- name (e.g., MM256),
- description (e.g., methane meter number 256),
- type name (e.g., methane meter),
- measured quantity (e.g., methane concentration),
- measurement unit (e.g., %CH4),

- name of a system that collects the data (e.g., THOR),
- range of measurements.

The *Source* table is described by means of *Location* dimension, that describes where in a coal mine it is located. The location has hierarchical structure, some sample hierarchy is presented in Fig. 3. The top-most level of hierarchy are formed by coal mine divisions. Divisions consist of seams, which are divided into mining areas. At the bottom of hierarchy there are mining workings.

The data warehouse is loaded with data by means of the ETL process designed for the main monitoring and dispatching systems for coal mining, which are deployed in Poland, Ukraine and China, e.g., THOR dispatching system [16] or Hestia natural hazards assessment system [6]. The ETL process was designed by means of Open Talend Studio [17].

During the tests of the created solution the data warehouse was loaded with 800 million records what resulted in 200 GB of data. Therefore, it enabled the performance tests and optimisation of both the logical data warehouse structure and database management system (PostgreSQL [18]). As a result the *Measurement* data table was partitioned according
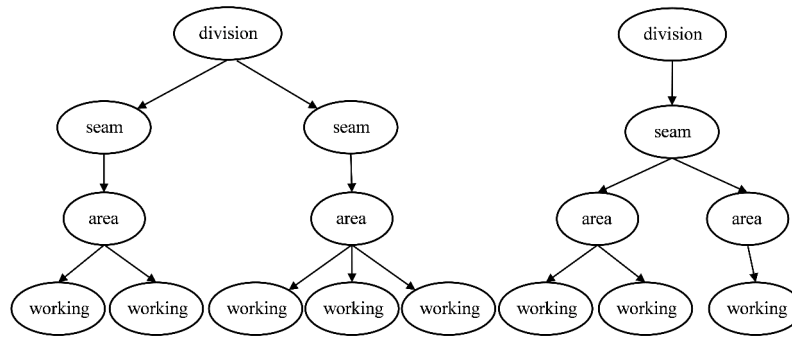
Fig. 3.   Location hierarchy in a coal mine

to the months of measurements and the indices for foreign keys in this table were created. On the DBMS side several configuration parameters were adjusted, e.g., shared_buffers, work_mem, maintenance_work_mem, checkpoint_segments, checkpoint_completion_target, effective_cache_size.

IV. Data preparation and cleaning

The goal of ETL2 module is to deliver integrated data (in a form of a uniform data set) coming from chosen sources (especially sensors) in a chosen time range.

The measurements can be collected with different frequencies. Additionally, some systems collect a new measurement only after significant (defined in a monitoring system) change of the measured value. Table II presents how the measurements of two methanometers can look like when collected directly from the data warehouse. The ETL2 process uniforms the data to the form where each recorded measurement represents the time period defined by a user, e.g., 1 second (Table III).

MN234	MN345	T[s]
0.1	0.1	0
0.2	-	1
-	0.2	4
0.5	?	7
0.3	0.3	9

TABLE II
Data collected directly from data warehouse (- means that the measurement value does not change, ? means a missing value)

MN234	MN345	T[s]
0.1	0.1	0
0.2	0.1	1
0.2	0.1	2
0.2	0.1	3
0.2	0.2	4
0.2	0.2	5
0.2	0.2	6
0.5	?	7
0.5	?	8
0.3	0.3	9

TABLE III
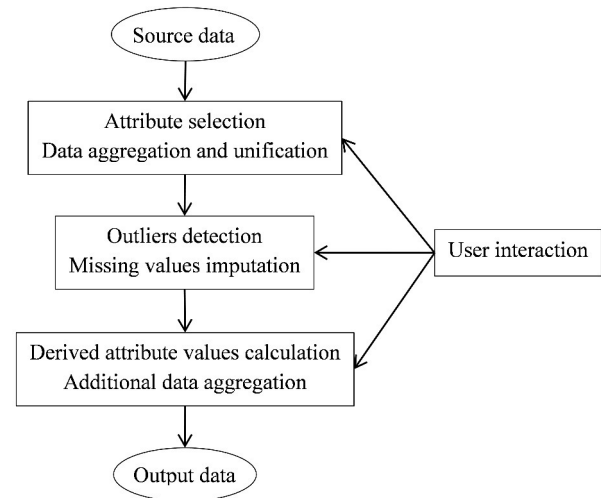Data prepared to the further transformation, cleaning, etc.

Fig. 4.   General characteristics of the data processing in ETL2 module

Within the ETL2 module there are also executed procedures of data cleaning, that identify outlier values and impute the missing values. This task is realised both by means of the simple functions presented below and by means of operators available in RapidMiner environment. Also data aggregation (e.g., 10 measurements are replaced with 1 measurement) and manual definition of derived variables (e.g., a new variable can be calculated as a sum of the values of two other variables) are performed by means of the methods included in ETL2 module.

The general scheme of data processing within ETL2 module is presented in Fig. 4.

As a result of the processing performed by means of the ETL2 module we receive a data set that can be either analysed (by means of analytical module), or utilised to prediction model creation (by means of prediction module), or utilised within diagnosis process (by means of expert system). All the phases of processing are performed as separate RapidMiner operators.

In order to select the variables that should be analysed a user can utilise THOR dispatching system 5, where each sensor

(and attributes) are presented on a map of the region of interest. The system that is being created enables in turn, data (time-series) visualisation in order to select the time periods, that are the most interesting from the analyst point of view. Fig. 6 presents the visualisation of time-series consisting of several thousands of records. The developed operator creating such visualisation utilises R environment.
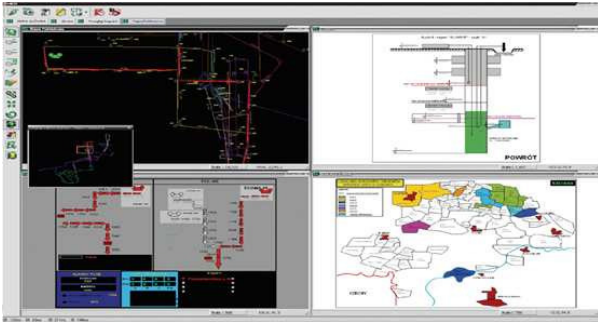

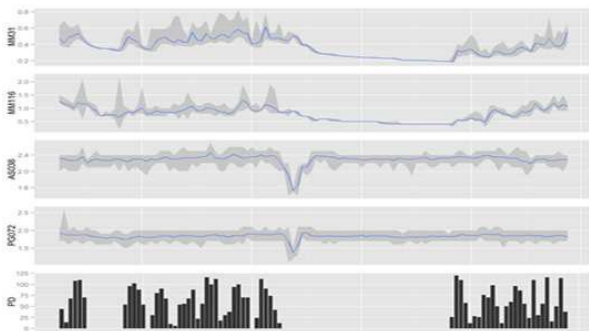Fig. 5.   Visualisation available in THOR dispatching system

Fig. 6.   Visualisation of exemplary time-series: methane concentration, air flow and mining cycle on a chosen longwall

Aggregation of the measurements replaces several values with a single one. The period of aggregation is chosen by a user, who sets a number of measurements that should be aggregated or a time unit defining the windows containing measurements to be aggregated. The following aggregation operators are available for each attribute: average, minimum, maximum, median, dominant, the number of occurrences. For each record being the result of the aggregation there is calculated a weight, that is inversely proportional to the number of missing values existing in the aggregated data. The weight calculation is also based on a weighted average for all the attributes. This approach enables us to reduce the number of missing values in data and introduce weights that can be utilised by the chosen methods (e.g., rule induction).

The operator that imputes missing values performs the analysis of each attribute separately. The following methods of changing the value or imputing the missing value can be utilised:

- a logical expression defining the replacing values (e.g., replace each value <1 with "low state"),

- the way how to receive the replacing values:
  - the value set by a user,
  - the last valid measurement,
  - average of the neighbouring measurements (with the parameter defining the number of neighbours),
  - linear regression of the two points (the last one before missing values section and the first one after this section),
  - linear regression of the data preceding missing values (with the parameter defining the window size).

The maximal number of consecutive missing values that can be imputed is defined as a separate parameter, as imputing the values for the long breaks in the measurements has no practical meaning. Therefore, the resulting data set can still contain missing values. In such case, the analyst can use a number of methods that are able to analyse data with missing values.

Introduction of a new derived variable can cover, among others, introduction of delays (the values of the previous measurements) or calculation of increments and trends (e.g., as an ordinal - increases, decreases). Another operator enables data smoothing by means of different filters (e.g., average, median). Finally, the last operator enables creation of dependent variable (decision variable). Typically, this variable contains the moved forward values of the chosen attribute, what enables to receive a proper prediction horizon. The operator defining the dependent variable has expanded functionality what enables e.g., to define the dependent variable as a maximal value of a given attribute in a defined time interval (e.g., 3 to 6 minutes in advance).

It is also important that within the developed framework the operators can be applied multiple times and in unrestricted order. Moreover, it is possible to pre-process data by means of the operators delivered by RapidMiner, that are dedicated to multidimensional analysis/identification of outliers and missing values (e.g., the operator applying local k-NN to missing values imputation).

When data pre-processing is finished, the whole process is saved according to RapidMiner-XML standard, that was created for the needs of the system. Thereby, the prediction module is able to transform the incoming data to the form that is acceptable by prediction models (see section V).

V. PREDICTION MODULE

Prediction module is based on, so called, prediction services. Prediction service is a webservice that predicts values of a variable (discreet or continuous) on the basis of input vector. Prediction service is inseparably connected with a model (regression or classification one) that is the basis of the prediction. The basic scenario of prediction service application is as follows:

1) Client sends a prediction execution request accompanied by a vector of conditional attributes and a timestamp.
2) Service calculates the prediction delivering the vector of conditional attributes as a model input. The attribute values come directly from the monitoring system, because

the data warehouse is not loaded online. The values of the attributes are transformed according to the dedicated ETL2 process to the form acceptable by the prediction model.

3) Service loads the results to a database.

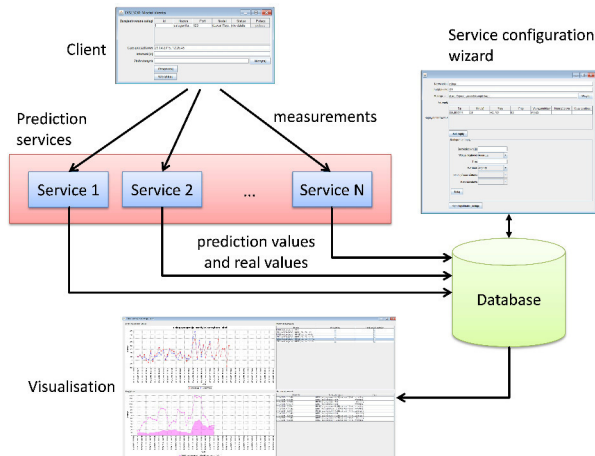The architecture of the prediction module is presented in Fig. 7.

Fig. 7. Architecture and operation of prediction module

Database, which is an internal RapidMiner repository, stores the description of a model and the transformations of the attributes. It also stores the information about training data, the parameters of the minimal model quality and both predicted and real values of dependent variable. Each model adaptation results in a new database entry what makes the history of the changes available to the users.

The predictions can be visualised and compared on a single plot with the real values that are measured. Such visualisation can be performed by a monitoring or dispatching system (e.g., THOR dispatching system), where predicted values are delivered as measurements of a virtual sensor and the values of both sensors (virtual and real) can be easly compared.

It is assumed for the current module version, that if the quality of the predictions decreases below a given threshold, then a new training set is automatically collected. The size of this new data set is the same as size of the original data. The model adaptation is performed by modifying only the parameters of the existing model (the method and algorithm is not changed). Next, the quality of the model is verified on the same data that triggered the model adaptation (these data are not the part of the new training data set). If the quality of the adapted model is satisfactory, then this new model is applied to prediction. Otherwise a message is generated stating that prediction cannot be continued and it is needed to come back to analytical module in order to create a new prediction model.

The configuration wizard enables to define the so-called quality monitoring rules. From the practical point of view there is no point in presenting the minimum model quality by means
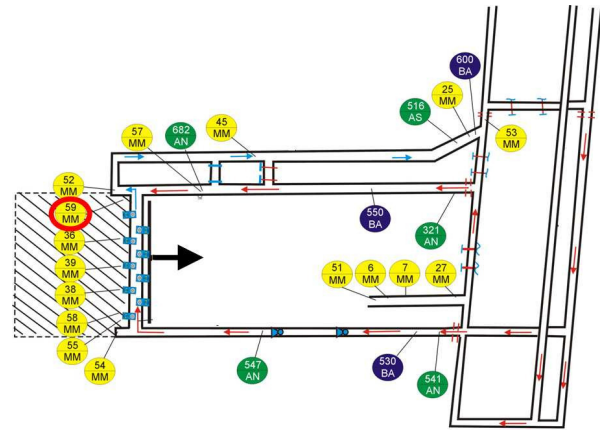

Fig. 8. Topology of the mining area and location of the sensors - MM59 sensor chosen as dependent variable is outlined a thick line
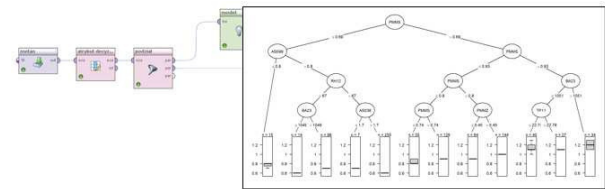

Fig. 9. The process of data preparation and prediction model creation together with the initial regression tree that was created

of the well known measures, such as overall classification accuracy, g-mean, specificity, sensitivity, RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), etc. Therefore, quality monitoring rules are based on: a sliding time-window (e.g., 1 hour) in which the quality is verified, frequency of the prediction calculation (e.g., 1 minute) and the indicators which are typically called *FalsePositive* and *FalseNegative*. The values of these indicators are explicitly defined by a user for each decision class or only for a target class, e.g., corresponding to "danger". Therefore, knowing the values of *FalsePositive* and *FalseNegative*, and a number of predictions that are calculated in a given time-window it is possible to calculate the values of almost all the possible quality measures of prediction model. In case of regression task the module allows so-called insensitivity, what means that the predictions that differ less than the given threshold from the real values are not treated as an error. Additionally, it is possible to define that the values within the given range (e.g., corresponding to the "normal" state) are not counted as errors.

VI. EXAMPLE OF THE SYSTEM APPLICATION TO THE TASK OF METHANE CONCENTRATION PREDICTION IN MINING EXCAVATION

The DISESOR system can be applied to several different tasks solving. This section presents an example, how the system can be utilised to methane concentration prediction.

Methane concentration monitoring is one of the main tasks of the natural hazard monitoring systems in mining industry.
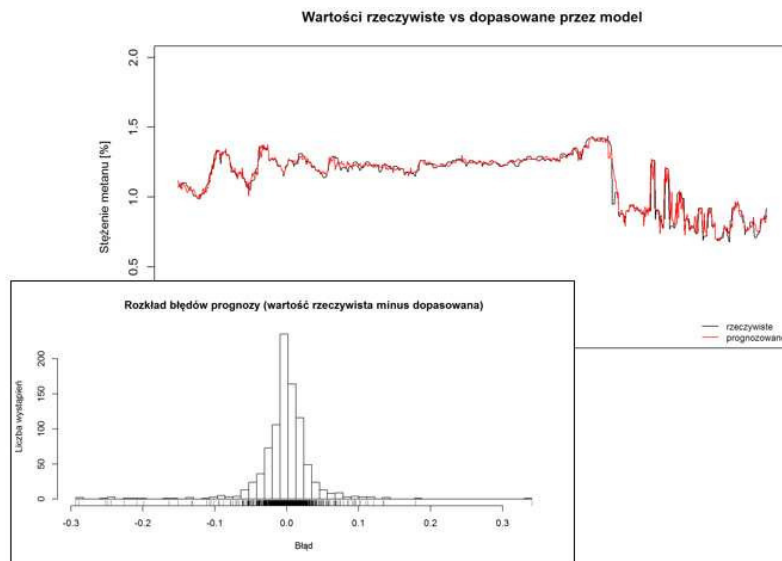
Fig. 10. The plot of the real methane concentration and the predicted maximum concentration together with the histogram of errors that are reported to a user

Such system is in charge of automatic and immediate shutdown of electricity within a given area, if a methane concentration exceeds a given alarm threshold. The power turn-on is possible after a certain time (from 15 minutes to even several hours), when the methane concentration decreases to the acceptable level. This results in large losses associated with downtime of production. Information from a soft (virtual) sensor presenting to a dispatcher the prediction of the methane concentration with a few minute horizon can allow the prevention electricity shut-down or can allow to lower the mining activity and increase the air flow if possible. Therefore, these actions allow to avoid undesirable situations and unnecessary downtimes.

The task of maximal methane concentration prediction with the horizon from 3 to 6 minutes was realised within the DISESOR system. By means of ETL2 module a set of the following sensors was selected: AN321, AN541, AN547, AN682, BA1000, BA603, BA613, BA623, MM11, MM21, MM25, MM31, MM36, MM38, MM39, MM41, MM45, MM52, MM53, MM54, MM55, MM57, MM58, MM59, MM61, MM81. The data were aggregated applying minimum operation to anemometer (AN) measurements, average operation to barometer (BA) measurements and maximum operation to methanometer (MM) measurements. The missing values were imputed applying linear regression method. As a dependent variable MM59 sensor was chosen. A map presenting the topology of the mining area and location of the sensors is presented in Fig. 8.

Analytical module is currently being developed and the analysis presented below is an example of the possible system apllication. Therefore, to create the examplary prediction model the method of regression tree induction was chosen arbitrary. The initial tree was created on the basis of data coming from 1 shift. The model and the list of sensors (variables) together with the defined transformations were forwarded to prediction model running a proper service. The time-window defined for prediction quality monitoring was set to 1 hour and the model adaptation was executed each hour regardless the minimum quality requirements. The adaptation could be executed more often if the minimum quality requirements were not met but there was no such situation. The data that were predicted were delivered on-line by the simulator of THOR system in order to simulate the real stream of measurements.

Fig. 9 presents the process of data preparation and the prediction model creation together with the initial regression tree that was created. Whereas, Fig. 10 presents the plot of the real methane concentration and the predicted maximum concentration together with the histogram of errors that are reported to a user. Currently, the user interface is in Polish as the deployment in Poland was planned in the project. However, the English and Chinese versions are also planned.

VII. CONCLUSIONS

The system that is being developed delivers the solutions for decision support of a dispatcher and process operator. This system is complete as it delivers the tools that can be applied to data storage, processing and preparation, and also to definition of the models based on expert knowledge (expert system) and the models based on the results of both historical and on-line data analysis. Due to the application and proper customisation of existing tools (RapidMiner, R) and development of the proprietary solutions (e.g., ETL2, rule induction and optimisation [19, 20], rough set operators [21] and semantic analysis of data mining processes [22] that are not available in RapidMiner) a user receives a broad set of tools that can be applied to different tasks. Finally, the case study that was presented shows that the system can be practically utilised in a coal mine industry.

REFERENCES

[1] M. Sikora and B. Sikora, "Rough natural hazards monitoring," in *Rough Sets: Selected Methods and Applications in Management and Engineering*. Springer, 2012, pp. 163–179. [Online]. Available: http://dx.doi.org/10.1007/978-1-4471-2760-4_10

[2] P. Kadlec, B. Gabrys, and S. Strandt, "Data-driven soft sensors in the process industry," *Computers & Chemical Engineering*, vol. 33, no. 4, pp. 795–814, 2009. doi: 10.1016/j.compchemeng.2008.12.012. [Online]. Available: http://dx.doi.org/10.1016/j.compchemeng.2008.12.012

[3] M. Sikora and B. Sikora, "Improving prediction models applied in systems monitoring natural hazards and machinery," *International Journal of Applied Mathematics and Computer Science*, vol. 22, no. 2, pp. 477–491, 2012. doi: 10.2478/v10006-012-0036-3. [Online]. Available: http://dx.doi.org/10.2478/v10006-012-0036-3

[4] A. Leśniak and Z. Isakow, "Space-time clustering of seismic events and hazard assessment in the zabrze-bielszowice coal mine, poland," *International Journal of Rock Mechanics and Mining Sciences*, vol. 46, no. 5, pp. 918–928, 2009. doi: 10.1016/j.ijrmms.2008.12.003. [Online]. Available: http://dx.doi.org/10.1016/j.ijrmms.2008.12.003

[5] J. Kabiesz, "Effect of the form of data on the quality of mine tremors hazard forecasting using neural networks," *Geotechnical & Geological Engineering*, vol. 24, no. 5, pp. 1131–1147, 2006. doi: 10.1007/s10706-005-1136-8. [Online]. Available: http://dx.doi.org/10.1007/s10706-005-1136-8

[6] J. Kabiesz, B. Sikora, M. Sikora, and Ł. Wróbel, "Application of rule-based models for seismic hazard prediction in coal mines," *ACTA MONTANISTICA SLOVACA*, vol. 18, no. 4, pp. 262–277, 2013.

[7] R. Kimball and M. Ross, *The data warehouse toolkit: the complete guide to dimensional modeling*. John Wiley & Sons, 2011.

[8] S. T. March and A. R. Hevner, "Integrated decision support systems: A data warehousing perspective," *Decision Support Systems*, vol. 43, no. 3, pp. 1031–1043, 2007. doi: 10.1016/j.dss.2005.05.029. [Online]. Available: http://dx.doi.org/10.1016/j.dss.2005.05.029

[9] M. Michalak, M. Sikora, and J. Sobczyk, "Analysis of the longwall conveyor chain based on a harmonic analysis," *Eksploatacja i Niezawodność - Maintenance and Reliability*, vol. 15, no. 4, pp. 332–333, 2013.

[10] M. Kalisch, P. Przystalka, and A. Timofiejczuk, "Application of selected classification schemes for fault diagnosis of actuator systems," in *Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on*. IEEE, 2014. doi: 10.15439/2014F158 pp. 1381–1390. [Online]. Available: http://dx.doi.org/10.15439/2014F158

[11] M. Grzegorowski, "Scaling of complex calculations over big data-sets," in *Active Media Technology*. Springer, 2014, pp. 73–84. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-09912-5_7

[12] E. Kozan and S. Q. Liu, "A demand-responsive decision support system for coal transportation," *Decision Support Systems*, vol. 54, no. 1, pp. 665–680, 2012. doi: 10.1016/j.dss.2012.08.012. [Online]. Available: http://dx.doi.org/10.1016/j.dss.2012.08.012

[13] RapidMiner. (2015) Rapidminer. [Online]. Available: http://rapidminer.com

[14] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2014. [Online]. Available: http://www.R-project.org

[15] A. Bifet, G. Holmes, R. Kirkby, and B. Pfahringer, "Moa: Massive online analysis," *The Journal of Machine Learning Research*, vol. 11, pp. 1601–1604, 2010.

[16] Sevitel. (2015) Thor. [Online]. Available: http://www.sevitel.pl/product,25,THOR.html

[17] Talend. (2015) Talend open studio. [Online]. Available: https://www.talend.com/products/talend-open-studio

[18] PostgreSQL. (2015) Postgresql. [Online]. Available: http://www.postgresql.org/

[19] T. Amin, I. Chikalov, M. Moshkov, and B. Zielosko, "Relationships between length and coverage of decision rules," *Fundam. Inform.*, vol. 129, no. 1-2, pp. 1–13, 2014. doi: 10.3233/FI-2014-956. [Online]. Available: http://dx.doi.org/10.3233/FI-2014-956

[20] U. Stanczyk, "Decision rule length as a basis for evaluation of attribute relevance," *Journal of Intelligent and Fuzzy Systems*, vol. 24, no. 3, pp. 429–445, 2013. doi: 10.3233/IFS-2012-0564. [Online]. Available: http://dx.doi.org/10.3233/IFS-2012-0564

[21] L. S. Riza, A. Janusz, C. Bergmeir, C. Cornelis, F. Herrera, D. Ślęzak, J. M. Benítez *et al.*, "Implementing algorithms of rough set theory and fuzzy rough set theory in the r package "roughsets"," *Information Sciences*, vol. 287, pp. 68–89, 2014. doi: 10.1016/j.ins.2014.07.029. [Online]. Available: http://dx.doi.org/10.1016/j.ins.2014.07.029

[22] A. Lawrynowicz and J. Potoniec, "Pattern based feature construction in semantic data mining," *Int. J. Semantic Web Inf. Syst.*, vol. 10, no. 1, pp. 27–65, 2014. doi: 10.4018/ijswis.2014010102. [Online]. Available: http://dx.doi.org/10.4018/ijswis.2014010102