APPLICATION OF RULE INDUCTION ALGORITHMS FOR ANALYSIS OF DATA COLLECTED BY SEISMIC HAZARD MONITORING SYSTEMS IN COAL MINES

ZASTOSOWANIE ALGORYTMÓW INDUKCJI REGUŁ DO ANALIZY DANYCH GROMADZONYCH PRZEZ SYSTEMY MONITOROWANIA ZAGROŻEŃ SEJSMICZNYCH W KOPALNIACH WĘGŁA KAMIENNEGO

The article presents the results of application of rule induction algorithms for predictive classification of states of rockburst hazard in a longwall. Used in mining practice computer system which is a source of valuable data was described at the beginning of this article. The rule induction algorithm and the way of improving classification accuracy were explained in the theoretical part. The results of analysis of data from two longwalls were presented in the experimental section.

Keywords: microseismic hazard prediction, classification, rule-based systems

W artykule przedstawiono możliwości zastosowania metody maszynowego uczenia, jaką jest indukcja reguł, do rozwiązania problemu predykcji stanu zagrożenia sejsmicznego w wyrobisku górniczym. Przedstawiono algorytm indukcji reguł, a także wyniki eksperymentów przeprowadzanych na danych pochodzących z dwóch ścian KWK Wesoła. Źródłem danych był system wspomagania stacji geofizyki górniczej Hestia. Przed przystąpieniem do analizy, dane zostały poddane agregacji tak, aby opisywały godzinowe i zmianowe przedziały czasu. W rozdziale czwartym umieszczono wyniki obrazujące dokładność proponowanych algorytmów w realizacji zadań predykcji zagrożeń sejsmicznych. Zadania predykcji definiowano w różny sposób (jako predykcję sumarycznej energii, jako predykcję zjawisk o określonej energii), różne były także rozważane horyzonty prognozy (zmianowy i godzinowy). Uzyskane wyniki wykazują, że przedstawiona metoda jest interesującą alternatywą dla innych metod opisanych w literaturze.

Słowa kluczowe: predykcja zagrożeń sejsmicznych, klasyfikacja, systemy regulowe

* R&D CENTRE EMAG, LEOPOLDA 31, 40-189 KATOWICE, POLAND
** INSTITUTE OF INFORMATICS, SILESIAN UNIVERSITY OF TECHNOLOGY, AKADEMICKA 16, 44-100 GLIWICE, POLAND
1. Introduction

One of the main tasks of geophysical stations in the coal mines is the estimation of current degree of rockburst hazard in active longwalls. In order to determine such degree, varied and detailed methods of hazard assessment are used, depending on the mine (most frequently the following are used: seismic, seismoacoustic and small-bore drilling methods). The final (complex) assessment is carried out on the basis of detailed methods and includes the assessments obtained by each of detailed methods as well as geological conditions prevailing in a given longwall (so-called method of mining reconnaissance). The description of detailed methods and complex method was published by the Central Mining Institute (Barański et al., 2007).

Due to the fact that the accuracy of rockburst hazard assessment methods usually applied in mining practice is far from perfect, which is proved by the big number of irrelevant assessments, one can observe constant development of new assessment and hazard prediction methods (Cianciara & Cianciara, 2006; Kabiesz, 2005; Kowalik, 1999; Dubiński et al., 1998; Kornowski, 2003a; Kornowski, 2003b; Rudajew & Číž, 1999; Sikora & Wróbel, 2009). To give an example, passive tomography method is the method which enables cyclical drawing of topographic maps of an interesting mining area, and on this basis, drawing conclusions about particularly threatened areas (Dubiński et al., 1998). The method of continuous prediction of rockburst hazard is the linear prediction method concerning prediction of total (seismic and seismoacoustic) energy, which will be emitted in the longwall at a given time horizon. The concept of linear prediction, presented i.a. in the works of Kornowski (2003a, 2003b), makes use of a mathematical apparatus implemented in the prediction of time series and enables hourly prediction of logarithmic sum of energy of events registered by the chosen geophone and energy of seismic events registered in a given longwall. On account of logarithmic energy value, minor prediction errors may result in higher values of the real prediction error. The method, except for the expecting energy value, provides also confidence intervals for the prediction. Another method of linear prediction is the indicating functions method (Cianciara A. & Cianciara B., 2006) basing on the probabilistic analysis of seismic hazard. In order to estimate the values of indicating functions the seismoacoustic emission is used (more precisely time intervals between events). The authors of this method claim that the value of indicating function constitutes the basis to assess the hazard and define time of tremors occurrence. The experimental tests show that in this method it is necessary to work out the way of translate the value of indicating function into hazard state assessment and the probability of tremors occurrence. The author of the method assumes that the increase of the value of indicating function shows hazard increase, alternatively the possibility of tremor occurrence. He has not defined, however, how the values and pace of indicating function increase translate into exact hazard assessment and the period of time when the assessment will be valid. Without the precise definition of those values, the method may cause too many so-called false alarms.

This article gives the opportunity to apply the method of machine learning understood as the induction of logical rules in order to create a classifier enabling the classification of two hazard states (“hazardous”, “non-hazardous”). Following i.a. the works of Kornowski (2003a, 2003b), we assume that the state in which the sum of seismoacoustic and seismic energy exceed in a given time period certain limit value will be considered as hazardous. The essential part of this research does not focus on the prediction of occurrence of a specific tremor but on the prediction, on the basis of gathered measurement data, the possibility of occurrence of a hazard situation in a given prediction horizon. As the threshold value of total (seismoacoustic (conventional) and seismic) energy separating the “non-hazardous” from “hazardous” states, the value $5 \times 10^5$J was chosen,
which is a more restrictive condition from the one presented in the works of Kornowski (2003a, 2003b). Good outcomes obtained for such defined prediction task resulted in undertaking an attempt to realize more difficult task concerning the prediction of seismic events of the energy more than 1*10^4 J.

An important aspect of our work is also the fact that the method proposed in this research may use all data available in the geophysical station, including the results of predictions generated by means of other (old or new) methods of hazard assessment.

This article is organized as follows. The second section describes the geophysical mining station supporting system, which is the source of data for the presented method of hazard assessment. Due to the fact that functionality and system architecture is complex, tasks and system possibilities were presented in a more detailed way. The third section presents the applied methodology of rule induction and classification mechanism of hazard states, the fourth section shows the results of experimental studies, finally, the fifth section includes the conclusions and directions of further work.

2. System supporting a geophysical mining station

The assessment of risk of rockburst by detailed and complex methods is connected with the necessity of aggregation and interpretation of measurement values registered by the devices in the mine underground. In order to automate interpretation process of gathered data in EMAG Centre, Hestia system was created (Sikora, 2003). It has been developed since several years and at the moment it has been implemented in several dozen geophysical stations both in Poland as well as in China, Russia and Ukraine.

The infrastructure of geophysical station of coal mine in which Hestia system is working has been presented in the first figure. Measurement data are usually obtained from two types of systems: seismic and seismoacoustic. In the first figure the representants of seismic systems are the systems marked as AramisE (EMAG Centre) and Multilok (Central Mining Institute), the representant of seismoacoustic systems is the system marked as AresE (EMAG Centre). The above mentioned systems gather measurement data and then after their adequate transformation, including i.a. localization of seismic events and computation of various types of aggregates (e.g. deviation of energy), transfer the processed information to the master system, which is Hestia system.

In the database of Hestia system the spatial and organizational structure of a mine is reflected (a mining plant, seam, area, longwall). For each longwall the information about the following is entered: hazard state resulting from the assessment of mining reconnaissance, longwall height, longwall type and the way of roof conducting. Except for measurement data transferred by specialist subsystems, database of the system stores information entered by the user in a manual way, such as information about roadway and longwalls advance as well as information about made drilling. The above mentioned set of information allows Hestia system to carry out the assessment of hazard state by means of a detailed and complex methods and to perform advanced queries to the database to create reports of various types.

Seismic method assessment (Barański et al., 2007) is carried out on the basis of: number and energy of registered seismic events, value of total energy of seismic events on five meters of longwall advance and the assessment of results of seismic events released in the longwall. Seismoacoustic method assessment (Barański et al., 2007) is carried out in AresE software and only hazard states (in the scale: a-lack of hazard, b, c, d – strong hazard), which are determined
for each geophone assigned to the assessed longwall, are transferred to Hestia system. According to an appropriate instruction (Barański et al., 2007) the influence on hazard assessment resulting from small-bore drilling method has the following aspects: longwall height, exploratory whole length, bore dust capacity and drilling run. Any information essential for assessment by means of small-bore drilling method must be entered by the system operator. On the basis of detailed assessments (the system also allows for data entry of assessment results from so-called supplementary methods) and the assessment resulting from the mining reconnaissance method, Hestia system using complex method determines the risk of rockbursts, which will be present in the longwall at the following shift. Assessment results and information summing up the risk of rockbursts in monitored longwalls are presented in a summary shift report.

Apart from summary report, Hestia enables the creation of any reports concerning the number, energy and types of registered seismic events (e.g. the report of total energy of seismic events). It is also possible to define filters limiting the number of information occurring in the report (e.g. minimal and maximal range of seismic events energy, events types etc.). Hestia gives also the opportunity to save chosen elements of the database to the popular file formats, i.a. *xls* and *txt*.

An important part of the activity of geophysical stations is the visualization of registered seismic events on the seam maps and by means of so-called histogram of tremors. Hestia Map software enables the creation of so-called histogram of seismic events and tremors as well as the visualization of seismic events on seam maps. Visualization of seismic events on the seam maps consists in drawing circles of a diameter appropriate to the energy of tremor and colour defining the type of an event. Map add-on to Hestia system is equipped with seam map editor. A library of graphic objects specific for seam maps (fault, abandoned workings, incline, seismometer, geophone, building, road etc.) constitutes a significant facilitation in map edition. In an automatic way, on the basis of entered longwall advance, so-called shading of abandoned workings is carried out – in other words, marking the already exploited region on the longwall area. Map add-on to Hestia system enables: to create seam maps (in the form of graphic layers);
to manage layers (i.a. projecting/hiding layers, map close-up); to create graphic reports (decade, monthly and quarterly maps drawn according to binding regulations (Barański et al., 2007), maps with seismic events of defined type and scope of energy; dynamic maps allowing to observe the sequence of events and longwall advance); to distance measurement between points; to visualize sensors (seismometers and geophones).

![Fig. 2. The section of map of mining area with visible location of seismic events and seismometers](image)

### 2.1. Preparing data for analysis

Due to data integration from different measurement systems, Hestia system may constitute a valuable source of data for different types of research on rockburst hazard assessment methods. In the carried out experiments the data gathered by Hestia system working in “Wesoła” Coal Mine were analysed. In this research two longwalls were considered, the data were aggregated in the hourly and shift periods. After data aggregation for each of the considered longwalls the following information was given:

- hazard assessment calculated by seismic method during shift (one of the values: a, b, c, d),
- hazard assessments calculated by seismoacoustic method during period (hour or shift) of data aggregation (one of the values: a, b, c, d),
- information about type of a shift (coal-getting or preparation shift),
• maximum total energy (expressed in conventional units) recorded by geophones, which monitor a given longwall, during period of data aggregation (geophone which records maximum energy is denoted by GMax to simplify further notation),
• maximum total number of pulses recorded by GMax geophone,
• deviation of total energy recorded by GMax geophone (the calculation of deviation was consistent with the existing seismoacoustic method (Barański et al., 2007)),
• deviation of a number of pulses recorded by GMax geophone (the calculation of deviation was consistent with the existing seismoacoustic method (Barański et al., 2007)),
• hazard assessment generated by seismoacoustic method for GMax geophone (one of the values: a, b, c, d),
• number of seismic events recorded during period of data aggregation,
• total energy of seismic events recorded during period of data aggregation,
• maximum energy of seismic events recorded during period of data aggregation.

If more than one geophone was assigned to a longwall then the set of variables contained also:
• average energy recorded by all geophones assigned to a given longwall during period of data aggregation,
• average number of pulses recorded by all geophones assigned to a given longwall,
• average deviation of energy recorded by all geophones,
• average deviation of a number of pulses recorded by all geophones.

A variable under prediction was the total seismoacoustic and seismic energy registered during aggregation period. In our experiments the prediction horizon amounted to one, meaning that we predict the energy with advance of one shift or one hour. The prediction may concern both the specific geophone and the whole longwall (then it may concern total, average or maximal value of seismoacoustic and seismic energy registered in the whole mine working). In order to omit the prediction of exact values for shift prediction, the scope of energy was divided into two intervals, below and above $5 \times 10^5$J. The values below $5 \times 10^5$J were considered as safety states (non-hazardous), the values above $5 \times 10^5$J were treated as dangerous states (hazardous). In the case of hourly prediction such restrictive setting of a hazard threshold caused that there were only few examples of “hazardous” state. So in the section of experiments concerning the hourly prediction, the threshold value was decreased up to $8 \times 10^4$J.

Due to the fact that the hazard assessment according to seismic method is usually carried out once in a shift, in the case of hourly aggregation, the values of variables, which contain the information about the assessments generated by seismic method, changed every eight records (every eight hours).

The detailed characteristic of the analysed dataset was presented in the fourth section.

3. Machine learning methods for the purposes of rockburst hazard prediction

3.1. Related works

Complexity of processes when the seismic tremors occur causes that the developed computational methods and statistic techniques turn out to be unsatisfactory to predict the hazard state
in coal mines, therefore, the usage of methods of computational intelligence for this purpose becomes more and more popular. Research studies on mining tremors most frequently use artificial neural networks (Łęski, 2008). In the majority of usage, the neural networks are used to predict the value of a variable associated with hazard assessment; this prediction is carried out on the basis of previously registered measurements. An example of such work is (Rudajev & Čiž, 1999) where the neural networks of various topologies where used for prediction of maximal cumulative amplitude and the number of seismic events. Prediction horizon lasted for a few days and prediction was done i.a. on the basis of longwall advance, number and maximal amplitude of registered seismic events in the previous periods. Numerous studies carried out by the Authors showed that the prediction of the number of tremors by means of neural networks gives better results than the prediction of energy of tremors. In the work of Kabiesz (2005) various types and networks topologies for prediction of tremors with energy greater than $10^4$ J were used. The prediction was made on the basis of longwall advance and the energy of seismic events registered in the periods preceding prediction. The author paid major attention to the selection of the form of input and output data as well as the topology of applied networks. The purpose of data transformation was to obtain the best possible prediction accuracy. In both of the mentioned studies it is underlined that in the research on hazard prediction it is necessary to take into consideration the additional information which could influence the increase of prediction accuracy. The drawback of neural networks usage is the difficulty to interpret obtained models, especially when the analysed data-set includes transformed variables (Kabiesz, 2005). Frequently, from the experts’ point of view concerning assessment of the obtained results not only the quality of prediction is important but also its premises, on the basis of which it was assigned.

In order to assess the hazard in mining longwalls so-called soft computing methods were also used (Kowalik, 1999). In the cited research work the application of fuzzy numbers was described in order to predict the expected, safe period of time, in which strong tremors should not occur. Similar research works were carried out using Markov chains.

The methods of computational intelligence were widely applied also in prediction of earthquakes. Also here the neural networks are often used (Bodri, 2001; Külahçı et al., 2008.). Another approach, similar to the one applied in this article is the application of rough set theory (Pawlak, 1991; Skowron & Rauszer, 1992) and methods of decision tree induction (Breiman et al., 1984) to predict the level of seismic activity on the basis of selected climate factors and concentration of radon in soil (Sikder & Munakata, 2007). The authors compare the results of predictions and the predictive models obtained by both methodologies showing experimentally that the obtained results are comparable and prediction accuracy is good.

Rule induction methods were also applied to solve the problem of the prediction of mining gas concentration in coal mines (Sikora M. & Sikora B., 2006; Sikora, et al., 2008) and the problem of diagnosis of devices (Sikora, 2007).

### 3.2. Induction of decision rules

Induction of decision rules is machine learning technique (Cichosz, 2000; Michie et al., 1994; Weiss & Kulikowski, 1991) realizing learning paradigm on the basis of examples. Induction of rules can be also applied in intensively developing computer science field which is the knowledge discovery in databases (Kubat et al., 1998). The rules as intuitive and easily interpreted depen-
ences are used for description and classification purposes. Decision rules (frequently known as classification rules) are usually represented in the form (1):

$$\text{IF } a_1 \in V_{a_1} \text{ AND... AND } a_k \in V_{a_k} \text{ THEN } d = v_d$$  \hspace{1cm} (1)

Rule induction is made on the basis of training dataset $\mathcal{DT} = (U, A \cup \{d\})$, where $U$ is a finite set of examples (also called objects or records) described by given set $A$ of features (also called conditional attributes) and the decision attribute $d$. Each attribute $a \in A$ is treated as a function $a: U \rightarrow D_a$, where $D_a$ is a domain of the attribute $a$. In accordance with this notation, in the rule (1) there is $\{a_1, \ldots, a_k\} \subseteq A$, $V_{a_i} \subseteq D_{a_i}$ and $v_d \in D_d$. The $a \in V$ expression is called conditional descriptor. The set of objects with the same value of decision attribute is called decision class (denoted as $X_v = \{x \in U; \ d(x) = v\}$).

Rules induction on the basis of training set can be done by means of various algorithms, in majority sequential covering algorithms are applied (An & Cercone, 2001; Grzymala-Busse, 1992; Sikora, 2006; Sikora, 2010), which create such a number of rules to assure that every object in the training set “is covered” by at least one rule. Other techniques are connected with the induction of all possible rules (Stefanowski & Vanderpooten, 2001) or induction of so-called minimal decision rules (Bazan et al., 2002, Sikora, 2010, Skowron & Rauszer, 1992). All algorithms use certain measures which decide either about the form of generated rules (i.e. which descriptors constitute its conditional part) or about that which of the inducted rules can be removed or joined.

### 3.3. Rule quality measures

Each rule (from given rule set $RUL$) in (1) form can be presented in the $\varphi \rightarrow \psi$ form. Any rule divides training set $U$ into two parts: $U_\varphi$ and $U_{\neg \varphi}$ which are determined, respectively, by antecedent $\varphi$ and consequent $\psi$ of the rule and therefore the set $U$ can be specified as $U = U_\varphi \cup U_{\neg \varphi}$ and $U = U_\psi \cup U_{\neg \psi}$.

The object $x \in U$ recognizes the rule (1) if and only if $\forall i \in \{1, \ldots, k\} \ a_i(x) \in V_{a_i}$. The object $x \in U$ supports the rule (1) if it recognizes it and $d(x) = v_d$.

The contingency matrix for rule $r \equiv \varphi \rightarrow \psi$ has the following form:

$$
\begin{array}{ccc}
  n_{\varphi \psi} & n_{\varphi \neg \psi} & n_{\neg \varphi} \\
  n_{\neg \varphi \psi} & n_{\neg \varphi \neg \psi} & n_{\neg \varphi} \\
   n_{\psi} & n_{\neg \psi} & n_{\varphi}
\end{array}
$$

where: $n_{\varphi} = n_{\varphi \psi} + n_{\varphi \neg \psi} = |U_\varphi|$ is the the number of objects which recognize the rule $\varphi \rightarrow \psi$; $n_{\neg \varphi} = n_{\neg \varphi \psi} + n_{\neg \varphi \neg \psi} = |U_{\neg \varphi}|$ is the number of objects which do not recognize the rule $\varphi \rightarrow \psi$; $n_{\varphi \psi} + n_{\neg \varphi \neg \psi} = |U_\psi|$ is the number of objects which belong to the decision class described by the rule $\varphi \rightarrow \psi$; $n_{\varphi \neg \psi} + n_{\neg \varphi \neg \psi} = |U_{\neg \varphi}|$ is the number of objects which do not belong to the decision class described by the rule $\varphi \rightarrow \psi$. The values $n_{\varphi \psi}, n_{\varphi \neg \psi}, n_{\neg \varphi \psi}$ and $n_{\neg \varphi \neg \psi}$ can be calculated on the basis of cardinalities of the sets $U_\varphi, U_{\neg \varphi}, U_\psi$ and $U_{\neg \psi}$; $n_{\varphi \psi} = |U_\varphi \cap U_\psi|$ (i.e. the the number of objects which support the rule $\varphi \rightarrow \psi$), $n_{\varphi \neg \psi} = |U_\varphi \cap U_{\neg \psi}|, n_{\neg \varphi \psi} = |U_{\neg \varphi} \cap U_\psi|, n_{\neg \varphi \neg \psi} = |U_{\neg \varphi} \cap U_{\neg \psi}|$.

There are two basic rule quality measures: accuracy (2) and coverage (3):

$$q^{acc}(\varphi \rightarrow \psi) = \frac{n_{\varphi \psi}}{n_{\varphi}} \hspace{1cm} (2)$$
Both measures considered simultaneously give the complete view of rule quality. According to induction by enumeration rule (Ajdukiewicz, 1974) it is assumed that the rules with high accuracy and coverage reflect the real dependences which are also true for objects beyond the analysed dataset. Obviously it is advisable to check whether the accuracy of induced rule is better than the accuracy resulting from the distribution of example between decision classes, in (Bruha, 1997; Guillet & Hamilton, 2007) one may approach a few proposals considering distribution of classes in the assessment of generated rules. It is easy to show that together with the increase of accuracy the rule coverage decreases, hence, big number of tests to define rule quality measures, which simultaneously take into account the accuracy and coverage of rules (An & Cercone, 2001; Bruha, 1997; Sikora, 2006; Sikora, 2010).

The analysis of ways of data acquisition, on the basis of which the rules applied in the process of hazard classification will be induced, shows that such data may be burdened with a considerable error, which results i.a. from inaccuracy of localization methods of seismic events, determining their energy and also human interference (e.g. higher activity of geophones may be caused by work of mechanical coal miner or nearby maintenance or repair work). Taking into consideration the origin of data in the carried out experiments two rule quality measures were used.

The first considered measures is the measure proposed by Cohen (4). It evaluates the strength of relationship between events “example \( u \) recognizes the rule” and “example \( u \) belongs to the decision class described by this rule” (which is reflected by difference \( nn_\varphi\psi – n_\varphi n_\psi \)) and between events “example \( u \) does not recognize the rule” and “example \( u \) does not belong to the decision class described by rule” (which is reflected by difference \( nn_{\neg\varphi\neg\psi} – n_{\neg\varphi} n_{\neg\psi} \)).

\[
q^\text{Cohen}(\varphi \to \psi) = \frac{nn_\varphi\psi + nn_{\neg\varphi\neg\psi} – n_\varphi n_\psi – n_{\neg\varphi} n_{\neg\psi}}{n^2 – n_\varphi n_\psi – n_{\neg\varphi} n_{\neg\psi}}
\]  

(4)

In a measure (4) denominator has the standardizing role. We may understand this measure as the function of two variables \( n_\varphi\psi, n_{\neg\varphi\neg\psi} \) (Fig. 3a). The range of variability of measure is the range \([-1,1]\) (from mathematical point of view it is better to talk about function than measure, but “measure” term was used on purpose in accordance with terminology) the higher value of the measure (4) the rule is considered as the better one.

The information theory, especially the concept of entropy, is used in works on optimal decision trees construction (Breiman et al., 1984; Cichosz, 2000). On the basis of this theory and information gain measures (i.e. \( \text{Gain} \) and \( \text{LimitedGain} \)) the following quality measure can be defined (5):

\[
q^\text{Gain}(\varphi \to \psi) = \text{Info}(U) – \text{Info}_{\varphi \to \psi}(U)
\]

(5)

where \( \text{Info}(U) \) is an entropy of the training example set (6):

\[
\text{Info}(U) = – \left( \frac{n_\varphi}{n} \log_2 \frac{n_\varphi}{n} + \frac{n_{\neg\varphi}}{n} \log_2 \frac{n_{\neg\varphi}}{n} \right)
\]

(6)

and \( \text{Info}_{\varphi \to \psi}(U) = (n_\varphi/|U|)\text{Info}(\varphi \to \psi) + ((|U|–n_\varphi)/|U|)\text{Info}(-\varphi \to \psi) \).
Entropy of a set of training examples is understood as the amount of information (in (6) formula it is the number of bits) necessary to communicate whether certain training object belongs or not to the decision class described by the rule \( \varphi \rightarrow \psi \). \( \text{Info}_{\varphi \rightarrow \psi}(U) \) value is the amount of information necessary to communicate whether certain object is or is not recognized by the rule \( \varphi \rightarrow \psi \). Gain measure is not monotonous because of the variables \( n_{\varphi \psi}, n_{\varphi \neg \psi} \) (Fig. 3b) and by this measure we can assess only the rules with accuracy higher than the accuracy of decision class resulting from the distribution of training examples between particular decision classes. Additionally, before evaluation of value of Gain measure it should be checked whether the investigated rule is accurate \( (n_{\varphi \neg \psi} = 0) \), if it is true then \( \text{Info}(\varphi \rightarrow \psi) = 0 \).

### 3.4. Rule induction algorithm

Induction (generation) of rules in the (1) form consists in selecting conditional attributes, which create conditional descriptors, and in determining their value ranges (i.e. the \( V_a \) sets). For given attribute \( a \in A \), the range of descriptor may have one of three forms:

- \( a = v \), where \( v \in D_a \),
- \( a \in V \), where \( V \subseteq D_a \) (e.g. seismic assessment \( \in \{a, b\} \) or number of pulses \( \in [3500, 4400] \)),
- \( a < v \) or \( a > v \), where \( v \in D_a \) (e.g. energy of seismic event > \( 1 \times 10^4 \)).

Below, the modified version (Sikora, 2006; Sikora, 2010) of MODLEM algorithm (Stefanowski, 1998) was briefly described. Modification allows for creating of so-called approximate rules, which may be inconsistent (inaccurate) with the set of training data. In the case of uncertain data (with errors, “noises” etc.) the approximate rules more easily capture general dependences occurring in the analysed dataset.

The MODLEM algorithm works in the following way; sorted in non-decreasing order values of each conditional attribute are one by one tested in order to find so called cut-off point \( g \). Cut-off point is in the middle, between two successive attribute \( a \) values (e.g. \( v_a < g < w_a \)) and
divides current range of values of attribute \( a \) into two ranges: \(( -\infty, g] \) and \([g, +\infty)\), and current set of training examples into two corresponding to these ranges subsets: \( U_1 \) and \( U_2 \). The cut-off point which minimize the value of expression (7) is optimal.

\[
\frac{|U_1|}{|U|} \text{Info}(U_1) + \frac{|U_2|}{|U|} \text{Info}(U_2)
\]  

(7)

The \( \text{Info}(U_i) \) denotes an entropy of set \( U_i \) (see formula (6)).

If the set \( U_1 \) contains more examples supporting generated rule than the set \( U_2 \), the range \(( -\infty, g] \) will be selected as conditional descriptor, otherwise the range \([g, +\infty)\) will be chosen. The descriptor is added to the previously created descriptors and together with them makes the conditional part of rule in the form of conjunction of conditions. If for the same attribute a two cut-off points \( g_1 \) and \( g_2 \) are chosen in the successive steps of algorithm, descriptor \([g_1, g_2]\) will be formed. If only one cut-off point \( g \) is found for some attribute \( a \) then descriptor \( a < g \) or \( a > g \) will be formed. If a cut-off point is not found for some attribute then this attribute will not appear in the rule.

For nominal attributes we do not look for cut-off points but search in a power set of a set of all possible values of this attribute. Nominal attributes are usually characterized by possessing a small number of different values, therefore, such a solution is characterized by acceptable computational cost (for example for the seismoacoustic assessment attribute we have four possible values \( a, c, b \) and \( d \), so we have to consider \( 2^4 - 1 \) subsets of the \( \{a, b, c, d\} \) set).

The algorithm constitutes the coverage of certain decision class, after generating each rule from the training set all the objects supporting the created rule are removed and the algorithm is applied to the rest of objects from the described class.

Generating accurate rules cause unfavourable (especially for noisy data) situation consists in the fact that the accurate rules are too fitted to training data, due to which some of them represent false dependences, which in an obvious way will influence both the quality of obtained classification and the gained knowledge. The modification of MODLEM algorithm (Sikora, 2006; Sikora, 2010) proposed by one of the authors of this work consists in application of rule quality measure for the evaluation of systematically created conditional part of rule. After addition (or modification) of subsequent conditional descriptor, the current form of the rule is assessed. Only the rule which obtained the highest quality value is selected as the output rule. The process of rule creation is ended when the addition of a new descriptor to the conditional part of rule causes the decrease of quality of the created rule.

In the modified version of MODLEM algorithm, criterion (7) decides about the sequence of descriptors’ addition and their value ranges, while the rule quality measure after adding each new descriptor or specializing the already existing descriptor checks whether the process of making rule more and more accurate can be stopped. As the works (Sikora 2006; Sikora, 2010) showed, the modified version of MODLEM algorithm allows to obtain better classification results of the unknown objects and to decrease the number of generated rules.

Regardless of the way of rule induction, the output rule set can be reduced by means of rule generalization or filtration algorithms (Sikora, 2010). Filtration works by removing from the rule set those rules which are irrelevant both for the description legibility and for the ability to generalize classifier.

A set of rules with identical conclusions in the form \( d = v \) is called a rule-based description of decision class \( X_v \).
Simple but efficient “Forward” filtration algorithm (Sikora, 2006; Sikora 2010) uses a ranking of rules created by any of the rule quality measures. The initial description of each decision class consists of one rule (the best rule), then one rule after another is added to description of each decision class; if the accuracy of decision class increases, the rule is left in the description, if not, another rule is considered. Ranking of rules decides about the order of rule processing and it is determined by rule quality measure (rules are sorted decreasingly towards the value of rule quality measure applied in the algorithm). Adding rule to the description of decision class is ended together with obtaining identical classification accuracy as non-filtered rule set or when all rules from input rule set are checked.

It should be strongly underlined that when filtration algorithm is working, classification is performed on either training set or so-called tuning set (usually tuning set and training set are disjoint), for the purposes of final verification of rule induction and filtration algorithms, the test set, independent from the training and tuning set, should be used.

### 3.5. Classifier and evaluation of its quality

The motivation of induction of decision rules is i.a. the fact that we want to use them (the rules) for prediction of value of decision attribute of the objects belonging both to the training table as well as the objects outside this table (so-called test objects). Assigning the value of decision attribute to the test object three cases may be encountered:

- none of the generated rules can recognize the test object,
- all rules recognizing the test object have identical conclusions,
- rules recognizing the test object have different conclusions.

In the first case, the classifier cannot take any decision, however, there are ways of solving this inconvenience, but they will not be applied in this article. In the surveys described in this paper, the test object which is not recognized by any of the induced rules is considered to be incorrectly classified object.

The problem at classification is the situation, in which the test object is recognized by the rules belonging to descriptions of different decision classes. What decision value should then be assigned to the test object? The answer to this question can be given in many ways. Most frequently, the solution to the occurring problem is found by voting mechanism. We can observe that for each of the generated rule we can calculate the value of one of the quality measures (2), (3), (4), (5), and then basing on works (Cichosz, 2000; Grzymała-Busse & Wang, 1996; Michie et al., 1994) we can:

- classify the object to this decision class, which is pointed at by the rule with maximal quality,
- sum up the quality of rules which recognize the test object and classify the object to this class for which this sum is the largest.

The way of proceeding allowing to generate the rules and also solve the conflicts during classification will be called a decision algorithm. The basic feature characterizing the efficiency of decision algorithm is its classification accuracy. In order to investigate the classification accuracy of a decision algorithm, the decision table available for the analysis is divided into two or three parts (Cichosz, 2000; Salzberg, 1997). The first one, on the basis of which the induction of rules is done is the training table, the second one, by means of which the efficiency of created
decision algorithm is investigated is the test table (sometimes the third part, called the tuning table is used). Training and tuning tables do not have to be disjoint, while the test table must be disjoint both from the training table as well as from the tuning table.

If \( DT_T \) is a test decision table and \( RUL \) is the set of rules then the classification accuracy rate of a decision algorithm making use of \( RUL \) set of rules and classifying the objects from \( DT_T \) is the value of expression (8):

\[
\text{accuracy}(RUL, DT_T) = \frac{|\{u \in U : f(u) = d(u)\}|}{|U|} \tag{8}
\]

In the (8) expression \( f \) is the function which assigns the \( u \) test object the values of decision attribute by means of \( RUL \) set of rules. The number in the numerator of the (8) expression is the number of correctly classified objects, i.e. such which value of decision function calculated by means of \( RUL \) set of rules is the same as the decision value resulting from the test table. When the number of examples representing individual decision classes is considerably different, the classification accuracy for each decision class \( X_v \subseteq U \) can be calculated separately (9):

\[
\text{accuracy}_{X_v}(RUL, DT_T) = \frac{|\{u \in X_v : f(u) = v\}|}{|X_v|} \tag{9}
\]

Depending on the way of division of input table into training and testing one we may obtain different classification accuracy. There are several methods which allow to determine a reliable estimator of the classification accuracy. Depending on the size of investigated data the most popular methods are the ones known under the following names: train-and-test and cross-validation (Cichosz, 2000; Michie et al., 1994; Salzberg, 1997).

**Train-and-test** method consists in dividing the analysed decision table in a random way into two subtables: training and testing. There are usually from 20% to 50% of all available objects in the test table. **Train-and-test** method is used when the analysed dataset contains more than 1000 objects. **Cross-validation** method estimates the value of the classification accuracy in a better way than **train-and-test** method. Due to computational complexity of rule induction algorithms, the method is used when the number of objects in decision table is smaller than 1000. In **cross-validation** method the decision table is divided in a random way in \( r \) disjoint and of the same cardinality subsets and then the \( r \) number of experiments is carried out. In each experiment, one of \( r \) subsets is a test set and the union of the remaining \( r-1 \) subsets is the training set (some may be a tuning set). After each experiment, the value of classification accuracy is computed. After carrying out \( r \) number of experiments, the final value of classification accuracy is computed as an arithmetic average of the accuracies from all \( r \) experiments. Usually \( r \) number is an integer from the range from 5 to 15.

4. Results of experiments

In order to verify the quality of classifiers obtained by methods described in the previous section, the shift and hourly data coming from two longwalls were analysed as the ones under the hazard of rockbursts. Below in the separate subsections the results obtained for prediction were presented. The prediction horizon lasted one shift and one hour respectively.
4.1. Hazard prediction in shift horizon

Before the analysis started, the data describing the conditions in two mining longwalls marked as SC508 and SC503 were taken from Hestia system. The number of examples and the distribution of examples between “hazardous” and “non-hazardous” decision classes were presented in the following way:

- SC508 - 864 records, 97 of which were assigned to “hazardous” state,
- SC503 - 1097 records, 188 of which were assigned to “hazardous” state.

In the analysed period of time, in 508 longwall, 563 seismic events of energy in class $10^3 J$, 82 events in class $10^4 J$ and 10 events of energy of the order of $10^5 J$ were registered. In the case of 503 longwall the numbers were 803, 128 and 6 respectively. As it was already mentioned in the second section, the threshold value of total seismoacoustic and seismic energy separating “hazardous” from “non-hazardous” state equalled to $5*10^5 J$. If more than one geophone was assigned to the longwall, in order to assess hazard state in prediction horizon which was of our interest, the energy of recorded seismic events and the energy (expressed in conventional units) recorded by the most active geophone were summed up, which is consistent with that proposed by Kornowski (2003a, 2003b).

It can be easily observed that the distribution of examples between individual decision classes is very uneven; in the described field of applications such a situation should be considered as normal because the “hazardous” states happen less frequently than “non-hazardous” states.

The first table contains the results obtained by the decision algorithm in which induction of rules was done by means of Cohen measure (the algorithm using the Gain measure obtained slightly worse results). After rule induction the rule filtration algorithm was applied, which also used the Cohen measure as the criterion of rule quality. In the rule filtration algorithm the whole training set was used as tuning set. The Cohen measure was also used during the classification of test objects. 5-fold cross-validation was used as testing methodology. Because more than one geophone was assigned to 508 longwall, the first table contains the results for decision table (denoted as SC508-avg) containing averaged values of measurements, as well as results for decision table (denoted as SC508-max) containing measurements recorded by the GMax geophone. 503 longwall was monitored only by one geophone (such data were available in Hestia system). The results included in the first table are the average results from 5-fold cross-validation method (the standard deviations were also given).

![Table 1](image)

**Classification results – test set – shift prediction horizon**

<table>
<thead>
<tr>
<th>Dataset (shift aggregation)</th>
<th>Accuracy (%) – „hazardous” state</th>
<th>Accuracy (%) – „non-hazardous” state</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC503</td>
<td>73.3 ± 0.00</td>
<td>93.6 ± 0.00</td>
</tr>
<tr>
<td>SC508-max</td>
<td>85.3 ± 0.01</td>
<td>71.9 ± 0.00</td>
</tr>
<tr>
<td>SC508-avg</td>
<td>70.1 ± 0.01</td>
<td>77.4 ± 0.01</td>
</tr>
<tr>
<td>SC503</td>
<td>74.4 ± 0.00</td>
<td>94.6 ± 0.00</td>
</tr>
<tr>
<td>SC508-max</td>
<td>84.6 ± 0.02</td>
<td>73.7 ± 0.00</td>
</tr>
<tr>
<td>SC508-avg</td>
<td>68.3 ± 0.05</td>
<td>79.5 ± 0.00</td>
</tr>
</tbody>
</table>
The analysis of the results presented in the first table leads to the conclusion that although unbalanced distribution of examples between decision classes, the classification accuracies of decision classes do not significantly differ from each other. It was obtained by means of appropriate rule quality measure. As we can see, rule filtration slightly influences the improvement of classification abilities; in particular, it contributes to the increase of accuracy of “non-hazardous” class. Filtration, however, has considerable influence on the limitation of the number of rules – the algorithm before filtration generates on average (regardless of datasets discussed here) 25 rules describing “non-hazardous” class and 3 rules describing “hazardous” class, after filtration the number of rules decreases and amounts on average to two rules for each decision class.

From the point of view of the worker of the geophysical station, it is important to predict precisely the “hazardous” state. However, the fact that the classifier should not in many cases predict the “hazardous” state in the case when it is “non-hazardous” is equally important. Too often and not accurate prediction of the “hazardous” state will cause the loss of trust to the classifier and “callousness” of the operator towards the warnings generated by the system. In a word, we want to cause high accuracy of both decision classes and make them similar. We can observe that in the case of data presented in the first table we deal with such a situation.

For comparison purposes with other methods, in the second table the classification results were presented. These results were obtained on the same datasets by means of RSES toolkit (Bazan et al., 2002) (LEM algorithm with rule shortening (Grzymala-Busse, 1992.) or exhaustive algorithm (Skowron & Rauszer, 1992) was used – depending on which of the algorithms obtained better results) and CART (Breiman et al., 1984) (gini index was used as the node split criterion), which are popular programs enabling the induction of decision rules.

### Table 2

<table>
<thead>
<tr>
<th>Dataset (shift aggregation)</th>
<th>Accuracy (%) – „hazardous” state</th>
<th>Accuracy (%) – „non-hazardous” state</th>
</tr>
</thead>
<tbody>
<tr>
<td>RSES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC503</td>
<td>49.4</td>
<td>97.1</td>
</tr>
<tr>
<td>SC508-max</td>
<td>46.4</td>
<td>93.0</td>
</tr>
<tr>
<td>SC508-avg</td>
<td>50.0</td>
<td>89.3</td>
</tr>
<tr>
<td>CART</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC503</td>
<td>87.7</td>
<td>86.8</td>
</tr>
<tr>
<td>SC508-max</td>
<td>81.4</td>
<td>73.9</td>
</tr>
<tr>
<td>SC508-avg</td>
<td>88.6</td>
<td>65.7</td>
</tr>
</tbody>
</table>

Comparing the results shown in the first and second table, we can observe that the results obtained by means of rule induction method presented in the third section and CART software are comparable. After calculating the average classification accuracy, it can be claimed that for prediction it is better to use the measurements from one geophone (that which shows the highest activity) than to apply the average results of the measurements. The rule induction method obtains here the average accuracy equal to 79.15%, CART 77.65%. For the averaged values of measurements the results are the following: rule induction 73.9%, CART 77.15%. RSES obtains much worse results of prediction of “hazardous” class.

Next series of experiments considers so-called start-up of the rule induction algorithm. In order to apply the algorithm in the new longwall we have to wait some time in order to use the data collected in this longwall by Hestia system for rule induction. The aim of the next experi-
The goal of the experiment was to check whether the rules generated on the basis of data from another longwall may be used for hazard classification in a given longwall. In the carried out experiment, the classifier was trained on the whole available dataset and applied to the whole available dataset coming from the second longwall. The results of the analysis were included in the third table.

<table>
<thead>
<tr>
<th>Datasets (shift aggregation)</th>
<th>Accuracy (%) – „hazardous” state</th>
<th>Accuracy (%) – „non-hazardous” state</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC503 as training set</td>
<td>53.6</td>
<td>91.3</td>
</tr>
<tr>
<td>SC508-max as test set</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC508-max as training set</td>
<td>76.0</td>
<td>89.5</td>
</tr>
<tr>
<td>SC503 as test set</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As we can see, the use of the classifier trained on the data from another longwall gives very good results in the case of prediction of “non-hazardous” state and satisfactory good results in the case of “hazardous” state. Better result for the hazardous state was obtained by training the classifier on the data coming from 508 longwall. The obtained results have considerable meaning for practical implementation of rule classifier in Hestia system. In the case of a new longwall, the hazardous state may be assessed at the initial period of classifier work according to rules already implemented in the system. Those rules will be the result of analysis of data coming from other longwalls. As the measuring data came from the monitored longwall another induction of rules may be considered, then the implemented set of rules may be replaced by new rules or new rules may constitute the supplement for already implemented rules in the system.

The last part of analysis of data describing consecutive shifts concerns the construction of induced rules. The analysis of rules describing “hazardous” state decision class shows that these rules contain the following attributes:

- energy recorded by GMax geophone,
- number of pulses recorded by GMax geophone,
- number of recorded seismic events with energy between $1 \times 10^3$ and $1 \times 10^4$,
- maximum energy of recorded seismic event.

It is an interesting observation that rules describing “non-hazardous” decision class contain the same attributes as listed above and additionally the following ones:

- deviation of the number of pulses (Barański et al., 2007) recorded by GMax geophone,
- deviation of total energy (Barański et al., 2007) recorded by GMax geophone,
- total energy of recorded seismic events.

It indicates that hazard assessments obtained by detailed methods (seismoacoustic and seismic methods) and complex method haven’t any influence on classifier predictions.

The example of two rules which are part of classifier obtained for longwall 503 is presented below (the energy recorded by geophone is expressed in conventional units (Barański et al., 2007) and it was highlighted by writing J” instead of J ).
IF energy of GMax geophone \( \leq 1.2 \times 10^5 \) J AND number of pulses of GMax geophone \( \in [2986,6932] \) AND total energy of seismic events \( < 5.7 \times 10^3 \) J THEN “non-hazardous” state (rule accuracy: 1.00)

IF energy of GMax geophone \( \geq 2.8 \times 10^5 \) J THEN „hazardous” state (rule accuracy: 0.72)

Taking into consideration the way of data aggregation (shift aggregation) and the prediction horizon (one shift), the first of these example rules should be interpreted as: “if during the current shift the energy (expressed in conventional units) recorded by GMax geophone is less than \( 1.2 \times 10^5 \) J and the number of pulses recorded by this geophone is in range \([2986,6932]\) and total energy of seismic events which occurred during this shift is less than \( 5.7 \times 10^3 \) J then the next shift will be non-hazardous”. Naturally, the word “non-hazardous” should be interpreted as it was defined in the task of hazard prediction (i.e. the sum of the conventional energy recorded by the most active geophone and the energy of seismic events will not exceed \( 5 \times 10^5 \) J during the next shift).

### 4.2. Hazard prediction in hourly horizon

Hazard prediction in hourly horizon required the decrease of cut-off value of energy, which separates the “non-hazardous” from “hazardous” states. Leaving not decreased threshold value (\( 5 \times 10^5 \) J) in both longwalls we obtain the small number of examples representing “hazardous” state – in the case of 503 longwall there are only three examples, in the case of 508 longwall we have 123 examples.

To increase the number of examples representing hazardous state, we decided to decrease the threshold value till \( 4.8 \times 10^4 \) J, citing the research work of Kornowski (2003b) it was agreed that such a solution is correct. The data from three summer months of 2007 were analysed, for 508 longwall 2203 records were obtained, 207 of which represented the “hazardous” state, for 503 longwall these numbers amounted to 2201 and 20 respectively. For 508 longwall, the records describing the “hazardous” state constitute about 10% of all records, for 503 longwall it was only one percent. Due to the bigger number of records, 3-fold cross-validation testing methodology was applied. In the case of hourly prediction, for 508 longwall as the conditional variables (attributes) both maximal and average (for bigger number of geophones) values of conventional energy and number of pulses were used.

### TABLE 4

Classification results – test set – hourly prediction horizon

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Accuracy (%) – „hazardous” state</th>
<th>Accuracy (%) – „non-hazardous” state</th>
</tr>
</thead>
<tbody>
<tr>
<td>The results obtained without the use of rule filtration algorithm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC503</td>
<td>76.1% ± 0.03</td>
<td>73.1% ± 0.02</td>
</tr>
<tr>
<td>SC508-max-avg</td>
<td>77.7% ± 0.00</td>
<td>72.0% ± 0.00</td>
</tr>
<tr>
<td>The results obtained with the use of rule filtration algorithm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC503</td>
<td>74.6% ± 0.00</td>
<td>70.4% ± 0.01</td>
</tr>
<tr>
<td>SC508-max-avg</td>
<td>81.1% ± 0.01</td>
<td>69.0% ± 0.00</td>
</tr>
</tbody>
</table>
TABLE 5

<table>
<thead>
<tr>
<th>Dataset (hourly aggregation)</th>
<th>Accuracy (%) – „hazardous” state</th>
<th>Accuracy (%) – „non-hazardous” state</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC503</td>
<td>60.0%</td>
<td>80.9%</td>
</tr>
<tr>
<td>SC508-max-avg</td>
<td>74.2%</td>
<td>72.1%</td>
</tr>
</tbody>
</table>

The results from the fourth table were obtained using Gain measure during induction of rules, rule filtration and classification process. The analysis of results contained in the table shows that for hourly horizon prediction, the accuracy of the “hazardous” state amounts on average to 75%, and the accuracy of the “non-hazardous” state is 71%. The results obtained by CART software (fifth table) are similar but are characterized by bigger difference in accuracy of decision classes.

Another experiment was to decrease the threshold which separates the “non-hazardous” from “hazardous” states. As the new threshold value 4.4*10^4J was appointed. Also in this case, the energy is the sum of conventional energy registered during one hour by most active geophone and the energies of registered seismic events during this hour. The decrease of threshold value was to increase the number of the „hazardous” states and check how it influences on classification accuracies of each decision class. The decrease of threshold value dividing the classes caused that for 508 longwall there were 942 examples of the “hazardous” state and for 503 longwall there were 232 such states. The results were presented in the sixth and seventh table.

TABLE 6

<table>
<thead>
<tr>
<th>Dataset (hourly aggregation)</th>
<th>Accuracy (%) – „hazardous” state</th>
<th>Accuracy (%) – „non-hazardous” state</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC503</td>
<td>82.9% ± 0.01</td>
<td>84.4% ± 0.00</td>
</tr>
<tr>
<td>SC508-max-śr</td>
<td>81.8% ± 0.01</td>
<td>72.0% ± 0.02</td>
</tr>
</tbody>
</table>

TABLE 7

<table>
<thead>
<tr>
<th>Dataset (hourly aggregation)</th>
<th>Accuracy (%) – „hazardous” state</th>
<th>Accuracy (%) – „non-hazardous” state</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC503</td>
<td>84.3%</td>
<td>81.8%</td>
</tr>
<tr>
<td>SC508-max-śr</td>
<td>74.0%</td>
<td>77.9%</td>
</tr>
</tbody>
</table>

For the decreased value of the threshold separating the “non-hazardous” from “hazardous” states considerably better classification accuracies of the both states were obtained; the reason behind it was better balance of the number of examples representing individual decision class. The average accuracy of rule induction methods is over one percent higher than the results obtained by CART software.

The last experiment with data describing hourly prediction horizon consists in considering the previous values of conditional attributes. New attributes were introduced into the decision
table. These attributes included the previous values of “basic” conditional attributes and they were called delayed variables. The idea of how to use delayed variables is based on the assumption that instead of using only aggregated values in the last hour, also the values from the previous periods of aggregation were used for prediction task. To give an example, having \textit{pulses} variable which represents the total number of pulses registered by the geophone at the last hour, to the prediction also the values of variables are used: \textit{pulses-1}, \textit{pulses-2}, ..., \textit{pulses-k}. The value of \textit{pulses-j} is the total number of pulses registered by the geophone at one hour but this value concerns the hour of registration which was carried out \(j\) hours ago. The results of the experiments were included in the eighth table, this time the higher values of the threshold (4.8*10^4J) were taken into account. The experiment was carried out for the longwall, in which more examples for the “hazardous” state were registered. The results in the eighth table were obtained by means of the rule induction algorithm without filtration (because as the results obtained by the algorithm without delays and filtration were considered as the point of reference – see table four).

<table>
<thead>
<tr>
<th>SC508-max-avg dataset (hourly aggregation)</th>
<th>Accuracy (%) – „hazardous” state</th>
<th>Accuracy (%) – „non-hazardous” state</th>
<th>Average classification accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without delays</td>
<td>77.7% ± 0.00</td>
<td>72.0% ± 0.00</td>
<td>74.85%</td>
</tr>
<tr>
<td>With the first delay</td>
<td>76.8% ± 0.01</td>
<td>70.9% ± 0.00</td>
<td>73.85%</td>
</tr>
<tr>
<td>Up to third delay</td>
<td>71.4% ± 0.01</td>
<td>75.7% ± 0.01</td>
<td>73.55%</td>
</tr>
<tr>
<td>Up to sixth delay</td>
<td>66.6% ± 0.00</td>
<td>78.3% ± 0.01</td>
<td>72.45%</td>
</tr>
<tr>
<td>Up to eighteenth delay</td>
<td>71.0% ± 0.00</td>
<td>79.0% ± 0.01</td>
<td>75.00%</td>
</tr>
</tbody>
</table>

The results included in the eighth table show that the use of delayed variables in such a form in which they were defined in this article does not improve the accuracy of predictions. Although the average classification accuracy for eighteenth delays is higher, this difference is not probably statistically important. Obviously the delayed variables may be defined in a different way, e.g. by summing up the number of pulses or energies from the following hours, counting the differences or increments etc., at this point such studies were not conducted. The results in the eighth table suggest that as the delayed variables are added, the accuracy of the “non-hazardous” class is increased and the accuracy of the “hazardous” class is decreased. Such a situation causes that the number of so-called false alarms will be smaller. These alarms consist in incorrect prediction of hazard (when there is no danger in reality), at the same time, however, the number of cases in which the system can detect the real hazard will be decreased. Certainly we may use ROC curve or gain criterion (Grzymała-Busse et al., 2005) in order to reach a compromise between so-called specificity and sensitivity of the algorithm.

### 4.3. Prediction of high-energy seismic events

The prediction of occurrence of a high-energy seismic event is a difficult task and considered by many as impossible to solve out. At this point we have to agree with such a thesis because any accurate methods of how to solve this problem have not been presented so far. Easier task is the prediction of occurrence in certain prediction horizon the seismic event of determined energy. As it was mentioned such works were described i.a. in (Kabiesz, 2005; Rudajev & Čiž, 1999).
Below in the ninth table the results of prediction of seismic event occurrence of \( \geq 1 \times 10^4J \) energy in the shift prediction horizon were presented. In the case of 508 longwall in the analysed period of time, there were 46 seismic events registered of energy of the order of \( 10^4J \) and 6 seismic events of energy of the order of \( 10^5J \). 508 longwall has the values of 112 and 6 respectively. The aim of the analysis was the prediction of occurrence of any of these seismic events at the time of approaching shift (the type of seismic events was not taken into account, e.g. which of them were rockbursts). When the experiment was carried out 5-fold cross-validation method was used.

<table>
<thead>
<tr>
<th>Dataset (shift aggregation)</th>
<th>Accuracy (%) – „seismic event occurs“</th>
<th>Accuracy (%) – “lack of seismic events”</th>
</tr>
</thead>
<tbody>
<tr>
<td>The results obtained without the use of rule filtration algorithm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC503</td>
<td>70.5% ± 0.03</td>
<td>68.0% ± 0.02</td>
</tr>
<tr>
<td>SC508-max</td>
<td>46.1% ± 0.08</td>
<td>71.2% ± 0.01</td>
</tr>
<tr>
<td>The results obtained with the use of rule filtration algorithm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC503</td>
<td>71.3% ± 0.05</td>
<td>68.7% ± 0.07</td>
</tr>
<tr>
<td>SC508-max</td>
<td>61.2% ± 0.06</td>
<td>66.0% ± 0.06</td>
</tr>
</tbody>
</table>

The quality of hazard predictions obtained in the ninth table is worse than the quality of shift prediction (for threshold of total energy equals to \( 5 \times 10^5J \); see table 1). The average accuracy of classification (by set of rules after filtration) amounts to 70% for 503 longwall and 63.4% for 508 longwall. The results included in the first table are for both longwalls better of about 15%, in the first table, however, the prediction task was put differently. It is worth seeing, that in the case of prediction of total energy, all those shifts were considered as “hazardous” in which the seismic events of energy were higher than \( 1 \times 10^4J \). Obviously, from the results presented in the 4.1 section, it is inferred that as hazardous were marked also some of those shifts during which the high-energy seismic events were absent. For 503 longwall there were in total 118 tremors of the energy higher than \( 1 \times 10^4J \) and there were 188 “hazardous” states for threshold value of total energy equalled \( 5 \times 10^5J \), for 508 longwall these numbers amounted 52 and 97 respectively.

5. Conclusion

The article presents the possibility of applying machine learning method (rule induction method) to solve the problem of prediction of seismic hazard in the mining workings. The rule induction algorithm was presented and also the results of experiments carried out on the data coming from two existing longwalls of “Wesola” Coal Mine. The source of data was geophysical station supporting system called Hestia, without which the collection of data necessary to conduct the experiments will be impossible. Before the analysis, the data were aggregated in such a way to describe hourly and shift time intervals. The fourth section included the results showing the accuracy of proposed algorithms in the realization of prediction tasks of seismic hazards. Prediction tasks were defined in different ways (as prediction of total energy and as
the prediction of seismic events of certain energy), the considered prediction horizons were also various (shift and hourly).

The obtained results show that the presented method is an interesting alternative for other methods described in the literature. The accuracy of methods which make use of neural networks or an apparatus of statistical mathematics is difficult to compare with the presented methods because:

- other authors as the accuracy criterion usually use the mean squared error; according to the authors of this work making use of average accuracy of classification gives better idea about the accuracy of used method;
- the analysed datasets coming from different longwalls; generally in the described field of usage there are not any benchmark datasets, which would be widely available and on which various methods could be compared.

Analysing obtained accuracies of classification the following conclusions can be drawn:

- shift prediction
  - the lowest accuracy of classification is obtained for the most difficult task which is the prediction of occurrence of seismic events of energy not lower than the defined threshold value (1*10^4J); the average accuracy of the method ranged from 64% to 70%;
  - better results were obtained for hazard prediction defined in a different way, in which the total energy of seismic events and conventional energy registered by the most active geophone were predicted; the average accuracy of the method ranged from 79.1% to 84.5%;

- hourly prediction
  - in the case of hourly prediction it was necessary to decrease the cut-off value separating the “non-hazardous” and “hazardous” states; for 4.8*10^4J threshold the average accuracy of classification from 72.5% to 75.5% was obtained; for the threshold lowered to 4.4*10^4J, these values amounted 76.9% and 83.6% respectively;
  - considering in the analysis the delayed variables do not considerably improve the accuracy of classification; this feature may result from the way of defining these variables and in the future work the way of building delayed variables will be changed (i.a. such aggregating functions as increments and sums will be used).

It is also worth paying attention to the fact that the importance lies also in parameterisation of rule induction algorithms and classification process. For shift prediction better results were obtained using the measure proposed by Cohen, for hourly data better results gave Gain measure.

Looking at the distribution of the number of examples describing “hazardous” and “non-hazardous” states and at even more unfavorable (unbalanced) distribution of the number of shifts in which the seismic events were registered of the energy higher than 1*10^4J, the obtained results should be considered as good, and in some cases even very good (especially for shift prediction of “hazardous” state). Among the carried out tests the biggest utilitarian meaning, in authors view, has the prediction of “hazardous” and “non-hazardous” states for threshold value of energy equals to 4.8*10^4J and the prediction of the occurrence of a seismic event of energy ≥1*10^4J.

The application of the method in the mining practice requires further intensive investigations. Currently the work over improving the accuracy of classification by the use of other accuracy measures and the use of fuzzy rules for prediction of hazard state is in progress. Various rule
quality measures are planned to be used in the rule induction algorithm for each decision class (other measures for “non-hazardous” state or “lack of high-energy seismic events” state) and other measures for the hazard state (or “the high-energy seismic event occurs” state).

From the point of view of considered field of applications the ideal solution will be also if the variables reflecting the predictions generated by other methods e.g. by the indicating functions method (it could be e.g. the value of risk function) or by the method of linear prediction (the estimated value of energy) were found among conditional attributes. Placing such features in a training dataset will require, however, the previous implementation of the chosen prognostic methods. It seems, however, that it is worth undertaking such a test. The characteristic feature which distinguishes the described methodology from other methods is the fact that during the analysis it takes into account the variables reflecting various aspects of conditions prevailing in a longwall. Certainly in the further research the variables reflecting i.a. the time intervals between the events and the pace of mining works (longwall advance) will be used as the new conditional attributes.

The characteristic feature of the presented analytical method is the fact that for the rule induction algorithm neither the type nor the semantic meaning of conditional attributes is important, the algorithm automatically chooses the features which have influence on the quality of predictions (these are the features which find their place in the premises of rules). The research carried out so far, which results were published in many works (Cianciara A. & Cianciara B., 2006; Kabiesz, 2005; Kowalik, 1999.; Kornowski, 2003a; Kornowski, 2003b; Rudajev & Čiž, 1999) usually took into account one aspect (one feature) of a conditions in the longwall (e.g. the seismoacoustic emission, time intervals between the events etc.), and the time series of the values of these features were analysed.

At the end it should be underlined that an important feature of continuously working classification system is also that this system should adapt to the changing environmental conditions and characterizes itself by high sensitivity and specificity. The worker of the geophysical station cares for exact and accurate perdition of the “hazardous” state, what is also important is the fact that the system should not in too many cases predict the “hazardous” state in the case when it is “non-hazardous”. Too often and not accurate prediction of the “hazardous” state will cause the loss of trust and “callousness” of the operator towards the warnings generated by the system. The continuously working system has to monitor constantly the quality of generated predictions and in the case of their worsening it should possess the abilities to choose a new training dataset (partially consisting of new data) and re-generating of rule set. Research works considering the adaptation abilities of the system will be, however, carried out in the second place, at present, works focus on further improvement of classification accuracy and analysis of data from other longwalls.

References


Received: 26 May 2009