Application of rule-based models for seismic hazard prediction in coal mines

Józef Kabiesz [1], ***Beata Sikora*** [2], ***Marek Sikora*** [3] *and* ***Łukasz Wróbel*** [4]

The paper presents results of application of a machine learning method, namely the induction of classification and regression rules, for seismic hazard prediction in coal mines. The main aim of this research was to verify if machine learning methods would be able to predict seismic hazard more accurately than methods routinely used in Polish coal mines on the basis of data gathered by monitoring systems. In this paper three classification and two regression tasks of prediction of seismic hazards in a longwall were defined. The first part of the paper describes the principles according to which the assessment of seismic hazard in Polish mines is made. These methods are called routine and allow to assess seismic hazard for a particular longwall. The next part of the paper discusses the algorithms of classification and regression rule induction and describes their use for seismic hazard assessment. The input data, which are the basis for rule induction, are: measurement data coming from seismometers and geophones, and the results of routine methods of hazard assessment. Conducted tests showed that automated hazard prediction based on induced rules gives better sensitivity and specificity of predictions than methods currently used in mining practice.

Keywords: *seismic hazard prediction, mining tremors, rule-based classification, rule-based regression*

Introduction

Mining activity was and is always connected with the dangers which are commonly called mining hazards. They are the causes of disasters and accidents, they are also an essential element in shaping industrial safety in coal mines. A special case of such a hazard is seismic hazard which frequently occurs in many underground mines. Seismic hazard is the hardest detectable and predictable of natural hazards and in this respect is comparable to an earthquake. Seismic activity and seismic hazard in underground coal mines occur in case of specific structure of geological deposit and the way of exploitation of coal. The number of factors, which influence on the nature of these hazards, is large and diverse, and relationships between these factors are very complex and insufficiently recognized. Such a situation, with a particularly strong intensity, occurs in the Upper Silesian Coal Basin where there are additional conditions connected with: multi-seam structure of deposit, consequences of the long history of exploitation of this area and complex surface infrastructure. In almost all mines of this area there are systems which detect and assess a current degree of seismic hazard.

One of the main tasks of coal mine geophysical stations is to determine the current state of seismic hazard (particularly, hazard of high-energy destructive tremor which may result in a rockburst) in underground mining places. Rockbursts, as a phenomena related with mining seismicity, pose a serious hazard to miners and can destroy longwalls and the equipment.

More and more advanced seismic and seismoacoustic monitoring systems allow a better understanding rock mass processes (Gale et al. 2001) and defining seismic hazard prediction methods (Gibowicz and Lasocki 2001). Accuracy of so far created methods is however far from perfect. These methods often require special, non standard measuring apparatus and that is the reason why some of them are not applied in a mining practice. New seismic hazard assessment and prediction methods, among others, are: probabilistic analysis (Lasocki 2005) which predicts the energy of future seismic tremors emitted in a given time horizon, linear prediction method (Kornowski 2003) which predicts aggregated energy (the sum of seismic and so-called seismoacoustic energy) in a given time horizon and so-called indicating function method which estimates the probability of strong tremor occurrence (Cianciara and Cianciara 2006).

Complexity of seismic processes and unbalanced distribution of positive ("hazardous state") and negative ("non-hazardous state") examples is a serious problem in seismic hazard prediction. Currently used statistical methods are still insufficient to achieve good sensitivity and specificity of predictions. Therefore, it is essential to search for new opportunities of better hazard prediction, also using machine learning methods. In seismic hazard assessment and prediction data clustering techniques may be applied (Kowalik (1999), Lasocki (2005)), and for prediction of seismic tremors artificial neural networks are used most often (Kabiesz (2005), Rudajev and Číž (1999), Leśniak and Isakow (2009)). In the majority of applications, the results obtained by mentioned methods are reported in the form of two

[1] *dr hab. inż. Józef Kabiesz*, Central Mining Institute, Plac Gwarków 1, 40-166 Katowice, Poland, jkabiesz@gig.eu

[2] *dr Beata Sikora*, Institute of Mathematics, Silesian University of Technology, Kaszubska 23, 44-100 Gliwice, Poland, beata.sikora@polsl.pl

[3] *Marek Sikora*, Institute of Informatics, Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland, marek.sikora@polsl.pl

[4] *mgr inż. Łukasz Wróbel*, R&D EMAG, Leopolda 31, 40-189 Katowice, Poland, lukasz.wrobel@emag.pl

states which are interpreted as "hazardous" and "non-hazardous". The fuzzy numbers (Kowalik 1999) and Markov chains are also used for hazard assessment defined in such a way.

Some of seismic hazard prediction methods outlined above were also applied in seismology for prediction of earthquake occurrence (Bodri 2001). Especially, rule induction and decision tree techniques were also applied for this purpose (Sikder and Munakada 2009).

The aim of this paper is to make an attempt at improving accuracy of seismic hazard prediction. Two hazard assessment methods, commonly applied in Polish hard coal mines, act as our reference point. These methods are so-called seismic method and seismoacoustic method which are described later in this paper. The methods presented in this paper use data routinely gathered by mine geophysical stations, what is important from the practical point of view. Moreover, rule-based data models can be also relatively easily interpreted by domain experts, and that, in turn, makes it easy to interpret results and to analyze the underlying causes of possible hazard. In the case of using artificial neural networks and repeatedly transformed data such an interpretation is not always possible (Kabiesz 2005).

This paper is organized as follows:

- second section discusses the genesis of tremors, the characteristics of seismic activity connected with longwall operation and relationship between tremors and hazard,
- third section presents methods of seismic hazard assessment used in Polish coal mines and effectiveness of these methods,
- fourth section briefly describes analyzed data sets and defines three classification and two regression tasks of seismic hazard prediction,
- fifth section discusses the methods of classification and regression rule induction applied to prediction of seismic hazard,
- sixth section gives a description of performed experiments,
- the final section provides a summary and suggestions for further work.

The characteristics of seismic activity associated with the exploitation of coal seams

The effect of coal seam exploitation is a creation of empty spaces in the rock mass. As the result, deforming rock strata put increased pressure on some of its areas. Zones of stress concentration are also formed in these strata. In the areas where the stress exceeds a critical value, rocks are destroyed causing emission of seismic tremors of different energy. The upper limit of energy of these tremors can exceed 10^9J. Sample graph of such emission expressed as the logarithm of the total energy E^C (the sum of energy of tremors E^W and energy of seismoacoustic emission E^{AE}) is presented on a logarithmic scale in Figure 1.

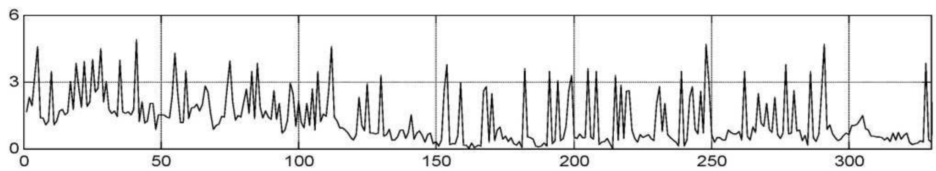


Fig. 1. Seismic emission $\log E^C$ *(the y-axis) during 350 hours (the x-axis) of work of longwall in one of the mines of Upper Silesian Coal Basin (Kornowski and Kurzeja 2008)*

The Gutenberg-Richter law (Gutenberg and Richter 1954) shows that the number of tremors increases exponentially with the decreasing value of their magnitudes. Characteristics of seismic activity described by equation (1), associated with the exploitation of many coal beds, can be more complex. These issues have been analyzed by many authors, for example Konopko (2009), (1984), (1994) presents typical forms of the Gutenberg-Richter law for Polish coal and copper ore mines.

The relationship between frequency of rockbursts occurrence and frequency of tremors occurrence was proposed by Konopko (1984). A formula was derived on the basis of analysis of seismic events which occurred between 1977 and 1983, and it was verified based on data from the years 1980-1999. On the basis of this analysis, the probability of rockbursts occurrence for mines located in the Upper Silesian Coal Basin can be defined by equation (1).

$$p = \log \frac{n_r}{n_t} = -14.598 e^{-0.3352 \log E} \tag{1}$$

where: the ratio of the number of rockbursts n_r to the number of tremors n_t expresses the probability of rockburst occurrence; E is the average energy of rockbursts in the given interval.

Equation (1) gives a relatively rough assessment of probability of rockburst occurrence, and the coefficients of equation (1) may vary depending on rock formations (Konopko 1994). In the paper by Bukowska (2006) a number of other factors having an effect on seismic hazard occurrence was proposed, among other factors, the occurrence of tremors with energies greater then 10^4J was listed.

This information in conjunction with geomechanical relationships between tremors and rockbursts justifies the use of sequence of recorded seismic activity for assessment and prediction of seismic hazard in the mines. The use of recorded seismic events for assessment and prediction is common practice and it is implemented not only in the Polish mining industry. The effectiveness of this method depends on the nature of processes of tremors generation, range, accuracy and reliability of the recording of seismic events, ability to analyze collected data, tools used for data analysis etc.

Classifications of seismic hazard in coal mines and their effectiveness

Two basic methods are routinely used for assessment and prediction of seismic hazard in Polish coal mines. These methods are called the *seismic method* and *seismoacoustic method* (Barański et al. 2007). Division of emission of seismic energy into these two categories in the Polish mining industry has historical roots connected with the type of measuring apparatus used for recording of geomechanical phenomena. This situation is gradually changing, including non-routine methods of seismic hazard prediction.

The seismic method

The essence of seismic method is the recording and analysis of tremors occurrence in mines. As Table 1 shows, relationships between tremors and rockbursts are foundations of the use of this method for seismic hazard assessment.

Observation of changes in the level of seismic activity and determination of increase (or decrease) of degree of hazard in comparison with previously determined degree is the essence of seismic assessment of seismic hazard. Depending on the intensity of seismic tremors occurrence, qualitative assessment (for low seismic activity) or quantitative assessment (for high seismic activity) is used for seismic hazard assessment. The level of seismic activity is determined on the basis of the number and energy of tremors recorded in the area of observed longwall in a certain time interval (a shift, day). Table 1 presents the basis of hazard assessment for quantitative method. Hazard assessment by the seismic method is made routinely every shift.

Tab. 1. Quantitative assessment of a seismic hazard according to observed seismic activity in the area of a longwall (Barański et al. 2007)

Rockburst hazard	Caved faces	Roadways
a No hazard	1. No tremors or single tremors with energies E of the order of 10^2J – 10^3J $E_{\max} \leq 10^4$J 2. $\sum E < 10^5$J per 5m of longwall advance	1. No tremors or single tremors with energies E of the order of 10^2J $E_{\max} \leq 10^3$J 2. $\sum E < 10^3$J per 5m of longwall advance
b Low hazard	1. Occurrence of tremors with energies E of the order of 10^2J – 10^5J $1 \cdot 10^4\text{J} < E_{\max} \leq 5 \cdot 10^5$J 2. $1 \cdot 10^5\text{J} \leq \sum E < 10^6$J per 5m of longwall advance	1. Occurrence of single tremors with energies E of the order of 10^2J – 10^3J $E_{\max} \leq 5 \cdot 10^3$J 2. $1 \cdot 10^3\text{J} \leq \sum E < 10^4$J per 5m of longwall advance
c Moderate hazard	1. Occurrence of tremors with energies E of the order of 10^2J – 10^6J $5 \cdot 10^5\text{J} < E_{\max} \leq 5 \cdot 10^6$J 2. $1 \cdot 10^6\text{J} \leq \sum E < 10^7$J per 5m of longwall advance	1. Occurrence of tremors with energies E of the order of 10^2J – 10^4J $5 \cdot 10^3\text{J} < E_{\max} \leq 5 \cdot 10^5$J 2. $1 \cdot 10^4\text{J} \leq \sum E < 10^5$J per 5m of longwall advance
d High hazard	1. Occurrence of tremors with energies E of the order of 10^2J – 10^6J $E_{\max} > 5 \cdot 10^6$J 2. $\sum E \geq 10^7$J per 5m of longwall advance	1. Occurrence of tremors with energies E of the order of 10^2J – 10^5J $E_{\max} > 10^5$J 2. $\sum E \geq 10^5$J per 5m of longwall advance

The seismoacoustic method

The essence of seismoacoustic method is recording and analysis of seismoacoustic emission (denoted by *AE*) occurring within a given longwall. Seismoacoustic emission is described by its intensity. The intensity concerns activity (that is, the number of registered events) or energy. Relationships between the seismoacoustic emission and seismic hazard are foundations of the use of the seismoacoustic method for seismic hazard assessment. In the seismoacoustic method the following factors are crucial for seismic hazard assessment:

- registration of the seismoacoustic emission,
- the number of pulses recorded by geophones, which is then converted by an appropriate formula for so-called conventional seismic energy.

The main criteria for assessment are changes in registered seismoacoustic activity and energy. Moreover, deviations (denoted in Table 2 by DEV) of values calculated during successive time intervals also influence defining one of the four states a, b, c, d of seismic hazard. Table 2 presents the way of defining the state of hazard which depends on the percentage changes in deviations DEV of the activity/energy values. Seismic hazard can be assessed by the seismoacoustic method once per shift or once per hour. In Table 2 there are presented the rules for one shift hazard assessment.

Detailed description of the hazard assessment algorithm is presented in (Barański at al., 2007).

Tab. 2. Defining the state of a seismic hazard using the seismoacoustic method on the basis of deviations (DEV %) as well as the direction of seismoacoustic activity/energy changes (Barański et al., 2007)

Time	$25 \leq$ DEV ≤ 100	$100 <$ DEV ≤ 200	Decrease of activity/energy after an increase of activity/energy such as $100 <$DEV≤ 200			DEV> 200	Decrease of activity/energy after an increase of activity/energy such as DEV>200		
			1 shift	2 shifts	>2shifts		1 shift	2 shifts	>2shifts
1 shift	a	b	a	a	current hazard state -1 after every 3 changes of activity/energy drop	c	c	c	current hazard state -1 after every 3 changes of activity/energy drop
2 shifts	a	c	b	b		d	d	d	
3 shifts	b	c	c	c		d	d	d	

Note: "current hazard state -1" means a reduction of the state of hazard by one degree (it remains unchanged in case of hazard state "a")

The effectiveness of seismic and seismoacoustic methods of hazard assessment

Knowledge of a current hazard state is a crucial element in management of production process and ensuring industrial safety. Therefore, accuracy and reliability of hazard assessment are important for mentioned above aspects of the mine work. However, seismic hazard assessment and prediction are very complex processes with a significant element of randomness.

At present, the majority of researchers claims that rockbursts which cause workers harm or the mine structure damage are impossible to predict. For example, depending on exploitation conditions, the recording of tremors with energy even up to 10^6J causes no undesirable effects in one case while in other headings phenomena of energy in the order of 10^5J can cause significant loss. For these reasons, many researchers variously define the hazardous state used in forecasting methods (Kabiesz (2005), Kornowski and Kurzeja (2008), Kornowski and Kurzeja (2012)). The seismic and seismoacoustic methods mentioned above are the result of the work of many domain experts. Although the quality of the predictions generated by both the methods is far from perfect.

The Polish Central Mining Institute analysis of selected cases of rockbursts shows that the seismic method predicted correctly the hazard state "d" in only about 17% of cases. If for the predictions of the state "d" the seismic and seismoacoustic methods are used at the same time (i.e. the state is projected to "d" while any of the methods indicates this state), such simultaneous use of both methods predicted correctly the hazard state "d" in about 20% of cases (Kabiesz 2010). In the aforementioned analysis, however, it is not considered in which number of cases possibility of rockburst occurrence was predicted, but in fact such a disaster did not take place. These so-called "false alarms" reduce the effectiveness of these methods and, what is more important, undermine the confidence of workers in the hazard monitoring system.

The method presented in this paper proposes a slightly different use of data collected by the mining geophysical stations then the seismic and seismoacoustic methods do. The model which is used for hazard assessment bases on archival data describing seismic activity in the area of a longwall. The basis of seismic and seismoacoustic methods are measurement data from geophones and seismometers, which are then processed according to algorithms realizing the hazard assessment described in Barański et al. (2007).

The method presented in this paper implements the scheme of so-called data exploratory analysis (so-called data mining), where a model of a phenomena is created only on the basis of an analysis of historical data. In principle, data mining methods do not analyze the physical causes of the phenomena, but try to create models as accurate as possible. In our case, this means the most accurate prediction of the state defined as "hazardous" while minimizing false prediction (see section 5.2). Of course, according to the usual definition of data mining (Weka) created data models may contain useful knowledge from the point of view of the field of application. This paper does not disregard this aspect completely (see section 6.3), but we focus mainly on the quality of the generated prediction and their comparison with the quality of predictions made using routine methods.

The application of this method poses some problems at the beginning of longwall exploitation, but after collecting the appropriate set of data, it allows to adjust better the way of hazard assessment to longwall conditions.

Data acquisition and data set description

Analyzed data are collected by the geophysical station supporting system called Hestia (Sikora and Mazik 2009) which has been introduced in several dozen Polish, Chinese and Ukrainian mines so far. This system assesses rockburst hazard on the basis of seismic and seismoacoustic methods (Barański et al. 2007), visualizes a mine on seam maps, standardizes and integrates data taken from seismic and seismoacoustic systems (Fig. 2). Hazard assessments are made individually for each longwall.

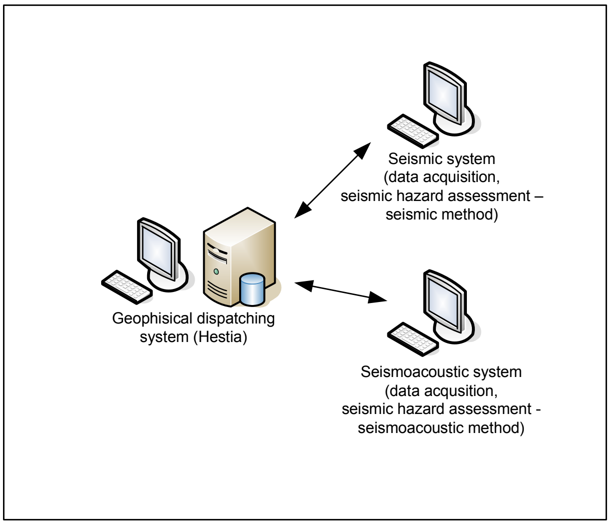


Fig. 2. The typical infrastructure of a geophysical station in a coal mine with the Hestia system installed

In the described studies two prediction horizons were considered: shift (eight-hour) prediction and one-hour prediction, but according to schedule of mining works eight-hour prediction is preferred. This means that the classifier makes the prediction of seismic hazard in one-shift advance.

The data acquisition is a continuous process, so raw data are aggregated before analysis. Aggregation process involves the replacement of a sequence of measurements recorded during the subsequent eight-hour intervals with one value. For example, aggregating measurement data collected during 100 shifts, we get a sequence of records (vector of variables) $(x_1, x_2, \ldots, x_{100})$, where x_t is a vector of aggregated measurement values describing the eight-hour time interval (i.e. one shift). For example, if we denote measurements recorded by the geophone G_1 by $g1$ and data aggregation is the summation of the number of pulses recorded by this geophone, than the variable G_1 stores the total number of pulses recorded during the subsequent shifts (thus, information on the total number of pulses registered by g_1 during the 10th shift we will get by referring to the component G_1 of the vector x_{10}).

During the study the aggregated values of the following variables were used:

1. hazard assessments made by seismic and seismoacoustic methods during a shift;
2. information about type of a shift (coal-getting or preparation shift);

3. maximum total energy recorded by geophones (which monitor a given longwall) during period of data aggregation (the geophone which records maximum energy is denoted by GMax to simplify further notation);
4. maximum total number of pulses recorded by GMax during period of data aggregation;
5. deviation of total energy recorded by GMax from the average energy recorded during eight previous periods of data aggregation;
6. deviation of a number of pulses recorded by GMax from an average number of pulses recorded during eight previous periods of data aggregation;
7. hazard assessment generated by seismoacoustic method for GMax (apart from the result of the seismoacoustic method generated for the whole considered excavation – see point 1 – the study used also the results of this method in the case of supplying the assessing algorithm only with the data from the geophone GMax);
8. number of seismic events recorded during period of data aggregation;
9. total energy of seismic events recorded during period of data aggregation;
10. 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;
- 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 recoded by all geophones.

Hazard assessment variables are symbolic and take values from the set {a, b, c, d}, where a stands for "no rockburst hazard", and d stands for high risk of rockburst. The rest of attributes are of the real type. Seismic hazard assessment using seismic method is performed routinely every 8 hours (i.e. every shift), therefore it changes every eight records for hourly aggregated data. This is due to the fact that in the case of the one-hour aggregation, any shift is described by 8 records.

The task of seismic hazard prediction may be defined in different ways (Bukowska (2006), Cianciara and Cianciara (2006), Kabiesz (2005), Kornowski and Kurzeja (2008), Kornowski and Kurzeja (2012)), but the main aim of all seismic hazard assessment methods is to predict (with given precision relating to time and date) of increased seismic activity which can cause a rockburst. In this paper we propose classification and regression approaches to this problem. The classification task is defined in three ways:

- first task – prediction of situations in which the sum of seismic energy of recorded tremors and energy recorded by GMax will exceed $5 \cdot 10^5$J in the next eight hours;
- second task – prediction of situations in which the sum of seismic energy of recorded tremors and energy recorded by GMax will exceed $4.8 \cdot 10^4$J in the next hour;
- third task – prediction that seismic tremor with energy greater than $1 \cdot 10^4$J will occur in the next eight hours.

The regression tasks are defined as follows:

- first task – prediction of sum of seismic energy of recorded tremors E^W and energy E^{AE} recorded by GMax which will be released in the next hour;
- second task – prediction of sum of seismic energy of recorded tremors E^W and energy E^{AE} recorded by GMax which will be released in the next eight hours.

Therefore, the presented tasks of hazard prediction base on the relationship between the energy of recorded tremors and seismoacoustic activity with the possibility of rockburst occurrence (as it was presented in Section 2 and 3). With the information about the possibility of hazardous situation occurrence, an appropriate supervision service can reduce a risk of rockburst (e.g. by concussion blasting) or withdraw workers from the threatened area. Good prediction of increased seismic activity is therefore a matter of great practical importance.

Building rule-based data models for seismic hazard prediction problem

Rule induction

In this study an algorithm which creates coverage of training data set and uses a greedy search strategy for creating elementary conditions of the rule premise (Fürnkranz 1999) was used for rule induction. Let us assume that there is a training set $DT = (U, A \cup \{d\})$, where U is a finite set of examples (also called objects or records) described by a given set A of features (also called conditional attributes) and the decision attribute d. Each attribute $a \in A$ is treated as a function $a\colon U \to D_a$, where D_a is a range of the attribute a.

If the purpose of rule induction is to classify the examples to one of several concepts (in our case, it will be the classification to one of the states "hazardous" or "non-hazardous"), then the values of a decision attribute are symbolic-type and indicate to which state each example from the training set belongs. If the purpose of the rule induction is to solve a certain regression problem, then for each training example, the values of the decision attribute indicate a specific numerical value (in our case, it will be the sum of seismic energy of recorded tremors and energy recorded by the GMax geophone).

The rule induction algorithm generates rules in the "**if** *Conditions* **then** $d = v$" form. In the case of the rules designed for classification tasks (so-called classification rules), the rule conclusion assigns the examples satisfying the rule condition to a particular concept. In the case of the rules designed for regression tasks (so called regression rules), the rule conclusion assigns a specific real value of the decision attribute to examples satisfying the rule premise.

The *Conditions* expression is a conjunction of the elementary conditions (i.e. w_1 and w_2 and . . . and w_n) which are also called conditional descriptors. Examples covered by the rule (i.e. satisfying all its elementary conditions w_i) are assigned to the decision class pointed by the rule. Each of the elementary conditions can be denoted as $a_i \ op \ Z_i$, where a_i is the conditional attribute, op is one of the relation symbols from the set $\{>, <, =, \in\}$ and Z_i is a numerical or nominal value from the range of the attribute a_i.

In the case of a classification rule with the conclusion $d = v$, the set of examples with the same value of the decision attribute is called the decision class (denoted by $X_v = \{x \in U : d(x) = v\}$). In the case of regression rules, the definition of decision class changes, because the value v located in the rule conclusion varies during the rule induction. For regression rules, a decision class is created of all training examples for which the decision attribute values are in the range $v \pm sd$. The sd denotes the standard deviation of decision attribute values for all examples covered by the rule.

Each rule generated can be presented in the $\varphi \to \psi$ form. Any rule $\varphi \to \psi$ divides the training set into two parts: U_φ and U_ψ which are determined, respectively, by the antecedent φ and consequent ψ of the rule, and therefore the set U can be specified as $U = U_\varphi \cup U_{\neg\varphi}$ and $U = U_\psi \cup U_{\neg\psi}$. For each rule $\varphi \to \psi$, the number of examples covered by this rule is defined as $n_\varphi = |U_\varphi \cap U_\psi| + |U_\varphi \cap U_{\neg\psi}|$, and the number of positive examples covered by this rule as $n_{\varphi\psi} = |U_\varphi \cap U_\psi|$. Let us assume additional denotation $n = |U|$ (for a number of examples in the training set) and $n_\psi = |U_\psi|$ (for a number of examples belonging to the decision class indicated by the rule conclusion, these are so-called positive examples). Based on these values many measures for rule quality evaluation as regards their accuracy (precision$(\varphi \to \psi) = n_{\varphi\psi}/(n_{\varphi\psi} + n_{\varphi\neg\psi})$) and coverage (coverage$(\varphi \to \psi) = n_{\varphi\psi}/n_\psi$) can be defined (An and Cercone (2001), Fürnkranz and Flach (2005), Sikora (2006), Sikora (2010), van Aswegen (2005), Yao (2003)).The process of rule premise generation consists in the selection of conditional attributes, which create conditional descriptors, and in determining ranges of their values. Our induction algorithm works in the following way: values of each conditional attribute sorted in non-decreasing order are one by one tested in order to find so-called cut-off point g. The cut-off point is in the middle, between two successive values of the attribute a (e.g. $v_a < g < w_a$) which separate positive examples from negative ones. The cut-off point g divides the current range of values of the attribute a into two ranges $\langle -\infty, g\rangle$ and $\langle g, +\infty)$, and current set of training examples into two subsets U_1 and U_2, corresponding to these ranges. The cut-off point which maximizes a value of the *RSS* (2) or *C*1 (3) measure, is optimal.

$$RSS(\varphi \to \psi) = \frac{n_{\varphi\psi}}{n_\psi} + \frac{n_{\varphi\neg\psi}}{n_{\neg\psi}} \tag{2}$$

$$C1(\varphi \to \psi) = \left(\frac{n_{\neg\psi} \cdot n_{\varphi\psi} - n_\psi \cdot n_{\varphi\neg\psi}}{n_{\neg\psi} \cdot n_\varphi}\right) \cdot \left(\frac{2 + \kappa(\varphi \to \psi)}{3}\right) \tag{3}$$

where $n_{\neg\psi} = n - n_\psi$ and $\kappa(\varphi \to \psi) = \dfrac{n \cdot precision(\varphi \to \psi) - n_\psi}{\left(\frac{n}{2}\right) \cdot \left(\frac{n_\varphi + n_\psi}{n_\varphi}\right) - n_\psi}$.

If the set U_1 contains more positive examples than the set U_2, the range $\langle -\infty, g\rangle$ will be selected as conditional descriptor, otherwise the range $\langle g, +\infty)$ will be chosen. The descriptor is added to the previously created descriptors and together with them creates the conditional part of rule in the form of conjunction of conditions. If a cut-off point is not found for some attribute a then this attribute will not appear in the rule. After adding the next conditional descriptor, the quality (i.e. the value of *RSS* or *C*1 measure) of the current form of the rule is evaluated. The rule with the best quality value is selected as the output rule. This phase of rule induction is called the rule growing.

In the next step of the rule induction the rule is pruned. The pruning process removes some of conditional descriptors from rules and also removes these of rules which are unnecessary in aspect of rule-based model optimization criterion. Rules are generated as long as they will cover all examples from the training set.

In the rule induction algorithm, various measures that evaluate the rule quality, and thus control the induction process can be used in the growing and pruning phases. In this paper, to simplify the description, we used the configuration of the algorithm in which in the growing and pruning phases the same measure is used. This is the *RSS* measure in the case of the classification rule induction, and the $C1$ measure in the case of the regression rule induction. This choice stems from our previous research on the rule induction algorithm efficiency (Sikora and Wróbel (2011), Sikora et al. (2012)).

The process of removing unnecessary rules, called also the rule filtration, is done by the Forward algorithm. The Forward algorithm, starting with one-rule descriptions of decision classes, builds a classifier, and then in each iteration it successively adds a rule from the ranking list to the rule-based model if adding of this rule increases the quality of the model. Rules sorted decreasingly according to a value of the used measure (2 or 3) create the ranking list. The detailed description of the Forward algorithm can be found in (Sikora (2006), Sikora (2010)). A detailed description of the described algorithm can be found in (Sikora (2006), Sikora and Wróbel (2011), Sikora et al. (2012)).

Quality and optimization of rule-based data models

Rule-based data models can be used to predict the value of the decision attribute for new, unknown examples. A test example can be covered simultaneously by many rules.

In this paper the rule voting mechanism is used during the classification process. The aim of the classification process is to assign an example (particularly a test example) $ts \in T$ to a corresponding decision class. If the example ts is covered by several rules which assign it to different decision classes, we deal with the ambiguous assignment problem. This ambiguity is resolved with a rule voting process. The confidence of each voting rule depends on its weighted Laplace (wLap) (4) value evaluated on the training set or so-called tuning set. For each decision class the sum of votes of rules covering the given example ts is calculated. The example ts is assigned to the class for which the sum of votes is the largest. As was shown empirically for a relatively large number of benchmark data, this approach gives the best quality of classification (Grzymała-Busse and Wang 1996). In the case of regression rules, the strategy known as *mean of conclusions* is used to determine a value of the decision variable. In this strategy, the expected value of the decision attribute is calculated as the average value of all conclusions of rules covering a test example.

$$wLap(\varphi \to \psi) = \frac{(n_{\varphi\psi} + 1) \cdot n}{(n_{\varphi} + 2) \cdot n_{\psi}} \tag{4}$$

The quality of obtained rule-based data model is usually measured by the classification accuracy or prediction error, which is calculated on a given data set independent of the training data set. The accuracy of classification may be, however, misleading in the case of unbalanced data distribution of decision classes (Fürnkranz and Flach (2005), Grzymała-Busse et al. (2005)). In this instance it is better to use sensitivity and specificity of the classifier. The classification accuracy for smaller decision class (called primary class or minority class) can be evaluated on the basis of so called classification confusion matrix (Fig. 3).

		actual class	
		primary	secondary
predicted class	primary	TP	FP
	secondary	FN	TN

Fig. 3. Classification confusion matrix

The all of correctly classified cases (examples) from the primary class are called true-positives (TP), incorrectly classified primary cases are called false-negatives (FN), correctly classified cases from the secondary class are called true-negatives (TN), and incorrectly classified secondary cases are called false-positives (FP). Three measures are usually used in the evaluation of the classifier quality: $sensitivity = TP/(TP+FN)$ (i.e. conditional probability of true-positives of a given primary class), $specificity = TN/(TN+FP)$ (i.e. conditional probability of true-negatives of a given secondary class) and *overall classification accuracy* $(TP+TN)/(TP+FP+FN+TN)$.

It is difficult to estimate what are the optimal values of *sensitivity* and *specificity*. High values of *sensitivity* and *specificity* are desirable for classification task but unfortunately there is usually a trade-off between each measure. This trade-off can be represented graphically by using a ROC curve (Fürnkranz and Flach 2005). The criterion which optimizes *sensitivity* and *specificity* of a classifier is called the *Class-gain* (5) measure (Bairagi and Suchindran (1989), Grzymała-Busse et al. (2005)).

$$Class\text{-}gain = Specificity + Sensitivity - 1. \tag{5}$$

Since it is not possible to determine the misclassification costs of positive examples in the seismic hazard prediction tasks, the measure (5) enables to find a good balance between capabilities of hazards identification (true-positives) and minimization of false alarms (false-positives). Evaluating quality of classifier by means of the measure (5) we are sure that the imbalanced distribution of positive and negative examples is taken into account.

In practice, if an alarm signal is generated by the classifier, it should be verified by a more detailed expert's analysis. The solution proposed in this paper seems to be a compromise between the most accurate prediction of real hazards and minimization of the number of situations identified as hazardous by classifier but recognized as safe after more detailed analysis by a geophysics.

In the case of regression problems solving, the *MAE* error (6) and the correlation coefficient between the predicted and actual values (7) are often used as a regression efficiency measures.

$$MAE = \frac{\sum_{i=1}^{n} |p_i - a_1|}{n} \tag{6}$$

$$Correlation = \sqrt{\frac{\sum_{i=1}^{n}(p_i - \overline{p}) \cdot (a_i - \overline{a})}{\sum_{i=1}^{n}(p_i - \overline{p}) \cdot \sum_{i=1}^{n}(a_i - \overline{a})}} \tag{7}$$

In formulas (6) and (7), n is the number of test examples, p_i are predicted values of the decision attribute of the test examples, a_i are the actual values of decision attribute of the test examples, $\overline{p}$ and $\overline{a}$ denote arithmetic means of all p_i and a_i.

Creating and applying rule-based data models

In practical applications creating and applying rule-based model consists of several steps: choosing the variables based on which elementary conditions will be created, preparation of the data set, selection a criteria evaluating the quality of created rules, choosing the technique of experimental evaluation of the rule quality and rule induction. Some of these steps can be carried out in an automatic way. The phase of data preparation must be – at least for the first time – supported by the user. If the resulting rule-based data model meets the minimum – also specified by the user – requirements of the quality, it is used for incoming data.

The quality of obtained rule-based data model is evaluated on a data set not used during the rule induction. Usually, in order to make the resulting score reliable the well-known technique called k-fold cross-validation technique is used [Weka]. The quality of the model is determined by one of the measures presented in Section 5.2.

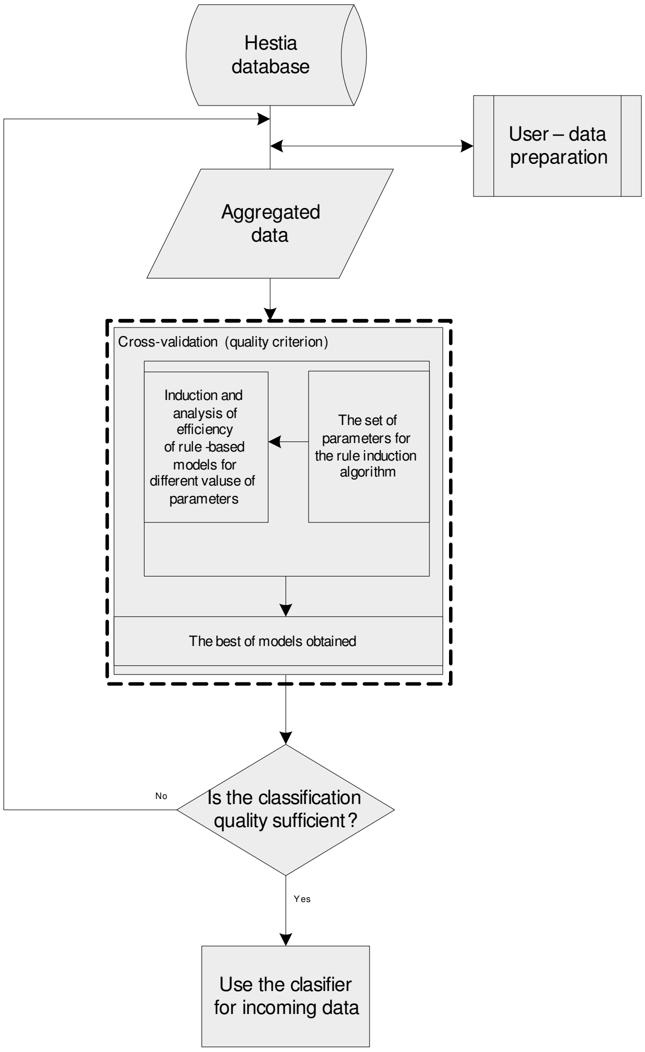
The described scheme of data analysis is shown in Figure 3. The first stage of the analysis, that is the selection of variables based on which the model will be constructed, the choice of data aggregation functions etc., requires user participation. The rest of the steps can be performed in an automatic way. During the rule induction there can be used various measures supervising the induction process (see Section 5.1), different induction algorithms can be used, too. The rule-based model that maximizes the value of user-defined quality criteria is chosen as the best one. In the case considered here it is the model that maximizes the Class-gain value.

The quality of the best model is confronted with the minimum quality requirements, defined by the user. In our case, there are compared the Class-gain values of the rule-based model and of routine methods. If the quality of predictions generated by the model is better than the quality of predictions resulting from the use of routine methods, the model can be applied to incoming data.

Experiments with data

Data obtained from two longwalls placed in the Mysłowice-Wesoła coal mine was used during the experimental tests. Both longwalls are threatened with rockbursts because of their geological structure.

The characteristic of analyzed data sets is presented in Table 3. As can be seen, all of these data sets are characterized by the unbalanced distribution of positive ("hazardous state") and negative ("non-hazardous state") examples. For example, for the longwall Sc508 and for one-hour seismic hazard prediction the examples which describe "hazardous state" account for about 10% of all of them, and for the longwall Sc503 which describe "hazardous state" account for only about 1% of all examples. Unbalanced data distribution is typical for classification of natural hazards because non-hazardous states occur more often than hazardous states. A similar problem also occurs in the case of hazard prediction of methane concentration (Sikora and Sikora 2006). In the experiments the filtration algorithm was executed in order to limit a number of rules pointing to "non-hazardous state" only, because the quantity of examples representing the decision class assigning "hazardous state" was definitely less than the number of examples pointing to "non-hazardous state". For the reason of differences in exploitation and geological conditions of both longwalls, for each longwall and for each prediction task the individual classifier was created.



Tab. 3. Characteristic of analyzed data sets

Longwall	No. of examples	No. of positives	No. of negatives
Eight-hour seismic hazard prediction (task 1 - prediction of total energy emission)			
Sc508	864	97	767
Sc503	1097	188	909
One-hour seismic hazard prediction (task 2 - prediction of total energy emission)			
Sc508	2203	207	1996
Sc503	2201	20	2181
Eight-hour seismic hazard prediction (task 3 - of high-energy tremors occurrence)			
Sc508	864	52	812
Sc503	1097	118	979

Results of hazardous state classification with the use of rule-based classifiers

Results of experiments are presented in Table 4. For eight-hour prediction a 5-fold cross-validation method was used. For one-hour prediction (where data sets are larger) a 3-fold cross-validation method was used. The table shows the average results obtained for independent test sets. In the successive columns the following values are given: name of the rule induction algorithm; the classification accuracy for "hazardous state" decision class; the classification accuracy for "non-hazardous state" decision class; Acc. – the overall classification accuracy and the Class-gain value. Each accuracy is expressed in percentage terms.

Tab. 4. Results of classification of seismic hazard states

Prediction task	Longwall	Algorithm	Acc. "hazardous state"	Acc. "non-hazardous state"	Acc.	Class-gain
1	Sc508	**RSS-rules**	77.1	77.8	77.8	0.55
		CART	81.4	73.9	74.8	0.55
	Sc503	RSS-rules	80.3	88.0	86.7	0.68
		CART	87.7	86.8	87.0	0.75
2	Sc508	**RSS-rules**	72.9	75.8	75.6	0.49
		CART	74.2	72.1	72.3	0.46
	Sc503	**RSS-rules**	65.9	75.9	75.8	0.42
		CART	60.0	80.9	80.7	0.41
3	Sc508	**RSS-rules**	67.1	67.3	67.3	0.34
		CART	88.4	42.8	45.6	0.31
	Sc503	**RSS-rules**	82.3	61.7	63.9	0.44
		CART	73.7	64.7	65.6	0.38

For comparison purposes the results obtained by the CART software designed for decision tree induction by the American Salford Systems company are also given.

The results presented in Table 4 show differences in classification accuracy between first and third prediction tasks. The same training sets were used for solving these tasks but differences in hazard definition result in difference in assigning examples to decision classes. As can be seen, the quality of classification depends on a way of "hazardous state" definition. It is clear that simultaneously with defining more difficult prediction tasks (ending with the prediction of the exact place and time of rockburst) the classification accuracy of methods presented in the paper as well as other methods will decrease. This fall in the classification accuracy can be also noticed in Table 6, which compares results obtained by the rule-based classifier with results obtained by methods currently used in Polish coal mines.

The sum of energy of tremors E^W and energy of seismoacoustic emission E^{AE} with the use of rule-based classifiers prediction results

The second part of the study focused on hourly and shift predicting of the sum of energy of tremors E^W and energy of seismoacoustic emission E^{AE} recorded by the geophone GMax. In the study, the logarithm value of $E^C + 1$ was predicted. As in the case of the classification task, for one-shift seismic hazard prediction the 5-fold cross-validation method was used. For one-hour seismic hazard prediction (where data sets are larger) the 3-fold cross-validation method was used. Besides the quality-driven rule induction algorithm, the algorithm RegEnder (Dembczyński et al. 2008) that applies the efficient boosting technique to construct rule-based data models was also used in the prediction process. In the RegEnder algorithm a number of induced rules has to be set by a user. In the experiments the RegEnder algorithm was run with the number of rules equal to 25, 50, 100, 200 and 300. Table 5 contains the best of the results (for the number of rules equal to 50). In this table, the columns named MAE and Correlation contain prediction errors and the correlation coefficient between the predicted and actual values of the sum of energy of tremors E^W and energy of emission seismoacoustic E^{AE} recorded by the geophone GMax, respectively.

The average absolute errors between the actual and the predicted values do not seem to be large. However, if we consider that we took logarithms of the values before prediction, the prediction accuracy is not good. In particular, for large values of energy e.g. 10^5J ($\log(100000) = 5$) errors equal to 0.3 bring large spread between the actual and predicted values of $\log(200000) = 5.3$. For the given example, we have therefore a huge difference in the prediction, amount to 10^5J. The prediction for the values of which we did not take logarithms leads to even worse results.

Tab. 5. The sum of the energy of tremors E^W and energy of seismoacoustic emission E^{AE} recorded by the geophone GMax prediction results

Prediction task	Longwall	Algorithm	MAE	Correlation
Shift predicting	Sc508	**C1-rules**	0.233	0.926
		RegEnder	0.244	0.898
	Sc503	C1-rules	0.336	0.824
		RegEnder	0.335	0.833
Hourly predicting	Sc508	**C1-rules**	0.201	0.713
		RegEnder	0.215	0.693
	Sc503	**C1-rules**	0.533	0.574
		RegEnder	0.561	0.581

The results obtained by the classifier (Section 6.1) and the results presented in this section show that the quality of the predictions decreases with specifying the prediction task. In particular, while specifying the energy that we want to predict.

First, the classifier achieves significantly better prediction accuracy than the regression model. Second, the best quality we note for hazard prediction task 1, and the worst for hazard prediction task 3 (see Table 4) and for the regression task.

Analysis of the obtained rule-based data models

The analysis of the obtained rule set will be started by checking which attributes occur in rules generated. In classification, rules describing the "hazardous state" decision class contain the following attributes:

- energy recorded by the GMax geophone;
- number of pulses recorded by the GMax geophone;
- deviation of the number of pulses recorded by GMax from the average number of pulses recorded during eight previous periods of data aggregation;
- number of recorded seismic events with energy between $1 \cdot 10^3$J and $1 \cdot 10^4$J;
- maximum energy of recorded seismic events.

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

- deviation of total energy recorded by GMax from the average energy recorded during eight previous periods of data aggregation;
- total energy of seismic events recorded during period of data aggregation;
- information whether a shift is coal-getting or not.

It is worth noticing that hazard assessments obtained by seismic and seismoacoustic methods occur in a few determined rules only. Therefore, they have no decisive influence on the assessment of hazard state made by the classifier.

In the regression rules, identification of the most frequent attributes is difficult. The number of induced rules exceeds the number of classification rules over five times. Moreover, almost all of the attributes describing the analyzed examples appear in the rules. A common feature of both the classification and regression rule sets induced is that the attributes indicating hazard results obtained by seismic and seismoacoustic methods appear only in a few rules.

The analysis of the rules obtained confirms the results of experiments described in the previous section. The classification rule-based data models solving the classification task are simpler and more legible than data models basing on regression rules. Therefore, in further part of this study we will focus on the classification models.

For the considered application domain it is also important whether predictions of created classifiers are better than predictions generated by routine methods in Polish coal mines. Routine methods assess rockburst hazard in four-scale a - no rockburst hazard, b - low hazard, c - moderate hazard, d - high hazard. The symbol $\geq \{b\}$ stands for any of the states b, c or d, consequently, $\geq \{c\}$ means any of the states c or d. In order to compare the accuracy of predictions, for the first task of seismic hazard prediction (prediction of the sum of seismic and seismoacoustic energy which can be emitted during the next eight hours) it is assumed that the hazard assessment of the state $\geq \{b\}$ generated by any of routine methods is equivalent to "hazardous state". For the second task of seismic hazard prediction (prediction of the sum of seismic and seismoacoustic energy which can be emitted during the next hour) it is assumed that the hazard assessment of the state $\geq \{b\}$ generated by the GMax geophone is equivalent to "hazardous state". For the task of prediction of high-energy seismic events occurrence it is assumed that the hazard assessment $\geq \{b\}$ generated by seismic method is equivalent to "hazardous state". More restrictive definition of hazard for routine methods, for example assumption that the hazard assessment $\geq \{c\}$ is equivalent to "hazardous state", worsens comparison results to disadvantage of routine methods.

The comparison between results obtained by the routine method and the rule-based classifier is shown in Table 6. The accuracy (expressed in percentage terms) of random assignment of examples to decision classes is also given in this table (this accuracy results from distribution of examples between decision classes). Analysis of these comparisons works to the rule-based classifier advantage.

In all presented cases the Class-gain value is definitely higher for the rule classifier than for routine methods. In all cases higher accuracy of "hazardous state" decision class was also noted for the classifier. However in three cases the classifier leads to generating higher number of false alarms than routine methods do. For these three cases a classifier in which no filtration was done after induction process was also trained (results for such type classifier were named "classifier wf"). As it can be seen in Table 6, such procedure causes decrease in accuracy of "hazardous state" with simultaneous increase in accuracy of "non-hazardous state". Results for the third prediction task and the longwall Sc503 are the exception. However it should be noticed that also in that case a value of the Class-gain criterion is higher for the rule classifier than for the routine method.

Tab. 6. Comparison of seismic hazard prediction effectiveness between routine methods and the rule-based classifier

Prediction task	Longwall	Method	Acc. "hazardous state"	Acc. "non-hazardous state"	Class-gain
1	Sc508	classifier	77.1	77.8	0.55
		routine	77.0	43.0	0.20
		a priori	11.2	88.7	0
	Sc503	classifier	80.3	88.0	0.68
		routine	56.8	57.0	0.14
		a priori	17.3	82.8	0
2	Sc508	classifier	72.9	75.8	0.49
		classifier wf	43.4	91.3	0.35
		routine	22.1	87.0	0.09
		a priori	9.3	90.6	0
	Sc503	classifier	65.9	75.9	0.42
		classifier wf	5.5	98.3	0.04
		routine	5.0	92.6	-0.03
		a priori	0.9	99.1	0
3	Sc508	classifier	67.1	67.3	0.34
		routine	48.3	47.8	-0.04
		a priori	6.0	93.9	0
	Sc503	classifier	82.3	61.7	0.44
		classifier wf	38.4	88.3	0.27
		routine	52.5	82.4	0.35
		a priori	10.7	89.2	0

Practical applications of seismic hazard classifier presented in this paper require appropriate training data sets for each longwall which should be included in the hazard monitoring system. Results presented in Table 4 and Table 6 use data collected for about 30% of a longwall mining period, this means that during the initial phase of longwall exploitation the hazard assessments have to base on existing hazard assessment methods. Table 7 presents the quality of classifiers which were trained with the use of smaller data sets. Training data accounted for about 10% of longwall mining period, the rest of data was used as the test set. Comparing results given in Table 4 and Table 7, it can be noticed that there are differences in the accuracy of decision classes and values of Class-gain between these two tables. The classifier trained on a smaller data set achieves results worse by 28% on average. These differences mean that during practical (routine) usage of the rule-based classification system, together with filling up the set of measurement data,

Tab. 7. Results of the test sets classification with the use of rule-based classifiers tested and optimized on smaller training and tuning sets

Prediction task	Longwall	Method	Acc. "hazardous state"	Acc. "non-hazardous state"	Class-gain
1	Sc508	classifier	61.2	83.6	0.45
	Sc503	classifier	87.2	78.0	0.65
2	Sc508	classifier	73.2	74.6	0.49
	Sc508	classifier wf	38.0	92.7	0.31
	Sc503	classifier	41.2	85.6	0.27
	Sc503	classifier wf	8.2	98.6	0.07
3	Sc508	classifier	40.8	75.0	0.16
	Sc503	classifier	62.4	72.5	0.35
	Sc503	classifier wf	31.6	88.9	0.21

the way of new classifier construction should be specified.

To automate the process of classifier building and determination of the classifier quality the following procedure is proposed. The measurement data set is divided into training and test parts. The method described in the fifth section is used for classifier building. Depending on the size of used data set, 5-fold cross-validation (less than 1000 records) or 3-fold cross-validation (more than 1000 records) should be used as a testing method. If an average value of the Class-gain measure of the classifier on test sets is greater than 0 and greater than an average value of Class-gain of routine methods, then it means that the rule-based classifier is worth using. Verifying whether a number of false alarms generated by the rule classifier is less than for routine methods can be an additional criterion for reasonableness of the classifier application. The system should also check the quality of own predictions at regular intervals and when results worsen (for example if they are worse than results of routine methods), the new rule set should be generated on the basis of new (including the newest data) data set.

Conclusions and future works

The paper has presented the potential use of the machine learning method, namely the induction of rules, for seismic hazard prediction in longwall mining. Three hazard prediction tasks related to real hazardous states in coal mines have been defined. Moreover, an attempt at hourly and shift predicting of the sum of energy of tremors E^W and energy of seismoacoustic emission E^{AE} recorded by the geophone GMax was undertaken. There has been no focus on predicting time of rockburst but, like in the case of other methods (Cianciara and Cianciara (2006), Kabiesz (2005), Kornowski (2003)), on identification of situations when increased seismic activity will occur within the area of monitored longwall.

The series of experiments have shown that the effectiveness of the presented way of rule-based classifier construction is greater than routine methods currently used in Polish coal mines. However, the regression task consisting in precise prediction of the sum of energy of tremors E^W and energy of seismoacoustic emission E^{AE} recorded by the geophone GMax failed.

Putting into practice the classification system for a longwall needs collecting historical measurement data coming from the longwall, training the classifier and checking whether it achieves better results than routine methods. Affirmative answer justifies the classifier application in practice.

On initial stage of a longwall exploitation, application of the presented method is impossible, because an appropriate data set based on which the prediction model can be created is not available. It seems to be impossible to create a general model suitable for seismic hazard assessment for any excavation. As it was presented in experimental part, the analysis of data coming from 10% of time of longwall exploitation is enough to get better results than in routine methods.

The hazard prediction method presented here can be used independently of assessments made by routine methods or by other currently developed hazard prediction methods. In the authors point of view, simultaneous use of several hazard assessment methods should lead to better seismic hazard identification, and thus contribute to better prevention of hazards of this type.

During further development of the classification system more conditional attributes such as longwall progress and time between successively recorded seismic events will be taken into account during the rule induction. Using assessments generated by other currently developed methods (linear prediction (Kornowski 2003), hazard functions method (Cianciara and Cianciara 2006)) as conditional attributes is also taken into consideration.

Acknowledgement: *This research was partly (to the second, third and fourth author) supported by Polish National Centre for Research and Development (NCBiR) grant PBS2/B9/20/2013 in frame of Applied Research Programmes.*

References

An, A., Cercone, N., 2001. Rule quality measures for rule induction systems - description and evaluation. Computational Intelligence 17, 409–424.

Bairagi, R., Suchindran, C., 1989. An estimation of the cut off point maximizing sum of sensitivity and specificity. Sankhya, Series B, Indian Journal of Statistics 51, 263–269.

Barański, A., Drzewiecki, J., Kabiesz, J., Konopko, W., Kornowski, J., Krzyżowski, A., Mutke, G., 2007. Rules of application of the comprehensive and detailed rockburst hazard assessment methods in hard-coal mines. volume 22. Central Mining Institute, Katowice, Poland. In Polish.

Bodri, B., 2001. A neural-network model for earthquake occurrence. Journal of Geodynamics 32, 289–310.

Bukowska, M., 2006. The probability of rockburst occurrence in the upper silesian coal basin area dependent on natural mining conditions. Journal of Mining Sciences 42, 570–577.

Cianciara, A., Cianciara, B., 2006. The meaning of seismoacoustic emission for estimation of time of the mining tremors occurrence. Archives of Mining Sciences 51, 563–575.

Dembczyński, K., Kotłowski, W., Słowiński, R., 2008. Solving regression by learning an ensemble of decision rules. Lecture Notes in Artificial Intelligence 5097, 533–544.

Fürnkranz, J., 1999. Separate-and-conquer rule learning. Artificial Intelligence Review 13, 3–54.

Fürnkranz, J., Flach, P., 2005. Roc 'n' rule learning - towards a better understanding of covering algorithms. Machine Learning 58, 39–77.

Gale, W., K.A.Heasley, Iannacchione, A., Swanson, P., Hatherly, P., King, A., 2001. Rock damage characterization from microseismic monitoring, in: Proceedings of 38th US symposium of rock mechanics, Washington DC. pp. 1313–20.

Gibowicz, S., Lasocki, S., 2001. Advances in geophysics. Academic Press, New York. volume 44. chapter Seismicity induced by mining: 10 years later. pp. 81–164.

Grzymała-Busse, J., Stefanowski, J., Wilk, S., 2005. A comparison of two approaches to data mining from imbalanced data. Journal of Intelligent Manufacturing 16, 65–573.

Grzymała-Busse, J., Wang, C., 1996. Classification methods in rule induction, in: Intelligent Information Systems, Proceedings of the Workshop, Deblin, Poland. pp. 120–126.

Gutenberg, B., Richter, C., 1954. Seismicity of the Earth and Associated Phenomena. Princeton University Press, New York. Frequency and energy of earthquakes 17-19.

Kabiesz, J., 2005. Effect of the form of data on the quality of mine tremors hazard forecasting using neural networks. Geotechnical and Geological Engineering 24, 1131–1147.

Kabiesz, J., 2010. Methods for assessment of rockburst hazard in coal mines' excavations. GIG, Katowice, Poland. volume 44. chapter The justification and objective to modify methods of forecasting the potential and assess the actual state of rockburst hazard. pp. 44–48. In Polish.

Konopko, W., 1984. Work Safety in Mining. Central Mining Institute, Katowice, Poland. volume 3. chapter A state and reasons of microseismic hazards in hard-coal mines in Upper Silesia. In Polish.

Konopko, W., 1994. Experimental basis for qualifying mining excavations in hard coal mines according to the rockbust hazard. volume 795. Central Mining Institute, Katowice, Poland. Research Report.

Konopko, W., 2009. Multisourceness of rock mass tremors, in: Conference proceedings - natural hazards in mining, Central Mining Institute, Katowice, Poland. pp. 97–103. In Polish.

Kornowski, J., 2003. Linear prediction of aggregated seismic and seismoacoustic energy emitted from a mining longwall. Acta Montana, Ser. A 22, 4–14.

Kornowski, J., Kurzeja, J., 2008. Short-term prediction of seismic hazards in mining. Central Mining Institute, Katowice, Poland. In Polish.

Kornowski, J., Kurzeja, J., 2012. Prediction of rockburst probability given seismic energy and factors defined by the expert method of hazard evaluation (mrg). Acta Geophysica 60, 472–486.

Kowalik, S., 1999. Prognosis of strong tremors in a mine with the application of fuzzy numbers, in: European Symposium on Intelligent Techniques ESIT'99, Orthodox Academy of Crete, Chania, Greece.

Lasocki, S., 2005. Probabilistic analysis of seismic hazard posed by mining induced events, in: Proceedings of sixth international symposium on rockburst and seismicity in mines, Australian Centre for Geomechanics, Western Australia. pp. 151–156.

Leśniak, A., Isakow, Z., 2009. Space-time clustering of seismic events and hazard assessment in the zabrze-bielszowice coal mine, poland. Int. Journal of Rock Mechanics and Mining Sciences 46, 918–928.

Rudajev, V., Číž, R., 1999. Estimation of mining tremor occurrence by using neural networks. Pure and Applied Geophysics 154, 57–72.

Sikder, I., Munakada, T., 2009. Application of rough set and decision tree for characterization of premonitory factors of low seismic activity. Expert Systems with Applications 36, 102–110.

Sikora, M., 2006. Lecture Notes in Artificial Intelligence. Springer-Verlag. volume 4259. chapter Rule quality measures in creation and reduction of data rules models. pp. 716–725.

Sikora, M., 2010. Transaction on Rough Sets XI, Lecture Notes on Computer Sciences. Springer-Verlag. volume 5946. chapter Decision rules based data models using TRS and NetTRS - methods and algorithms. pp. 130–160.

Sikora, M., Mazik, P., 2009. Towards the better assessment of a seismic hazard - the hestia and hestia map systems. Mechanization and Automation of Mining 3, 5–12.

Sikora, M., Sikora, B., 2006. Application of machine learning for prediction a methane concentration in a coal-mine. Archives of Mining Sciences 51, 475–492.

Sikora, M., Skowron, A., Wróbel, L., 2012. Rule quality measure-based induction of unordered sets of regression rules. Lecture Notes in Computer Science 7557, 162–171.

Sikora, M., Wróbel, L., 2011. Data-driven adaptive selection of rule quality measures for improving the rule induction algorithm. Lecture Notes in Computer Science 6743, 278–285.

van Aswegen, G., 2005. Routine seismic hazard assessment in some south african mines, in: Proceedings of sixth international symposium on rockburst and seismicity in mines, Australian Centre for Geomechanics, Western Australia. pp. 437–444.

Yao, Y., 2003. Entropy Measures, Maximum Entropy and Emerging Applications. Springer, Berlin. chapter Information-theoretic measures for knowledge discovery and data mining. pp. 115–136.