Safety Science 50 (2012) 918–922

Contents lists available at SciVerse ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/ssci

Data mining applications in evaluating mine ventilation system

Jianwei Cheng [a,*], Shengqiang Yang [b]

[a] Department of Mining Engineering, West Virginia University, Morgantwon, WV 26506, USA
[b] State Key of Laboratory of Mine Resource and Safety Exploition, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China

ARTICLE INFO

Article history:
Available online 27 August 2011

Keywords:
Mine ventilation
Date mining
Early warning model
Rough set
Support Vector Machine

ABSTRACT

A mine's ventilation system is an important component of an underground mining system. It provides a sufficient quantity of air to maintain suitable working environment. Therefore, the status of mine ventilation should be tracked and monitored as a timely matter. Based on former findings and in-depth analysis of mine ventilation systems, a proper early warning model is proposed in this paper for such considerations to improve the mine ventilation safety. The model itself is comprised of two sub-models, and two data mining techniques are used to assist in building each sub-model. One is the optimal indexes selection model which applies the Rough Set theory (RS) to assist the selection of best ventilation indexes. The other is the risk evaluation model based on the Support Vector Machine (SVM) to classify the risk ranks for the mine ventilation system. Testing cases have been used to demonstrate the applicability of this integrated model.

1. Introduction

A mine's ventilation system is an important component of an underground mining system. It should provide a sufficient quantity of air to the underground mine workings, to dilute methane and other contaminants, maintain suitable working environment and prevent accidents from happening. Very often, ventilation is a limiting factor for coal mine production (Cheng et al., 2010).

In the running stage of mine ventilation, its status could not be kept constant, and may change timely due to production requirements, mining laws and regulation, etc. Generally, the coal mine ventilation is a super complicated system. Lots of influence factors could control or impact the behaviors of system. Thus, stating from the quantitative point, the system shows a fluctuant wave around a certain value. However, if this wave swing is too large to control by the system itself, it may indicate that any potential risks existing.

Many methods or models were proposed to evaluate or assess the mine ventilation system. Jalali et al. (2009) proposed their definition of ventilation network evaluation. In this model, the most important factor to influence the network running was considered to be the branch air resistance. Therefore, by using the conventional methods to determine reliability in the transportation or electrical network, the reliability of each branch in the network was defined. Wang (2004) set up an evaluating model for the mine ventilation system reliability based on a BP neural network approach. A number of influence factors associated with the system reliability were considered in his model in order to achieve the most scientific results. Cheng (2008) and Cheng et al. (2010) introduced an integrated comprehensive method for selecting and evaluating the most suitable mine ventilation system. Using this method, the severe influence caused by anthropogenic factors or single index can be avoided in the final result. Thus, the selection and evaluation procedure of this method would ensure the selected mine ventilation system is the more rational and economical. Hatakeyama et al. (1992) analyzed the dynamic mine ventilation using airflow rate based on anemometer measurements. It could provide a real-time understanding the ventilation conditions, and also was expected to calculate the CH_4 or CO gas emission from surrounding strata to minimize the risk of disaster. Mitchell (1996) summarized accident prevention strategies to avoid loss of the property or life during a mine fire event. However, most proposed methods failed to identify the relationship between the historical ventilation records and the potential risk.

In this paper, an integrated early warning model is proposed to improve the mine ventilation safety due to such above considerations. The model itself is comprised of two sub-models. One is called the optimal selection model and the other is the risk evaluation model. All of them are based on the data mining technique which is an outstanding tool to mining the connections between the historical data and the induced risks. A test-case demonstration result shows that this integrated model has good applicability and could be applied in practices.

2. Optimal mine ventilation indexes selection model

2.1. Rough Set theory

The rough sets (RS) theory introduced by Pawlak (1982, 1991) has often proved to be an excellent mathematical tool for the

* Corresponding author. Tel.: +1 304 293 7680.
E-mail address: jcheng@mix.wvu.edu (J. Cheng).

0925-7535/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ssci.2011.08.003

analysis of a vague description of objects (Greco et al., 2001). It has good abilities to investigate the incomplete, inconsistent, or fuzzy information system, and then based on its special analytical procedure and inference method, to discover the implicit knowledge or underlying rules existing in the disordered data themselves. Recently years, the RS theory has been applied as a new tool into the data mining. It shows good performances in the areas of artificial intelligence, knowledge development, pattern recognition, and classification and so on.

2.2. Attribute reduction algorithm based on Rough Set

An interesting question in RS is whether there are attributes in the information system framework which are more important to the knowledge represented in the equivalence class structure than other attributes. Often, we wonder whether there is a subset of attributes which can, by itself, fully characterize the knowledge in the system. The process of achieving this aim is called the attribute reduction. The procedure is shown in the following sections:

2.2.1. Setting up the information system framework

Let $I = (U, A)$ be an information system, where U is a non-empty set of finite objects (the universe) and A is a non-empty, finite set of attributes such that $a : U \to V_a$ for every $a \in A$. V_a is the set of values that attribute may take. The information system assigns a value $a(x)$ from V_a to each attribute and object X in the universe.

2.2.2. Establishing the divisive matrix of decision table

The divisive matrix was firstly proposed by Showron (Zhang et al., 2001). It can easily express the in-distinguished relationships within the complicated information system, and help to assist in identifying the "core" of the system. The element in the divisive matrix can be defined as the following equation:

$$a(x,y) = \left| \begin{array}{ll} \varnothing & D(x_i) = D(x_j) \\ a \in A| & f(x,a) \neq f(y,a) \end{array} \right. \tag{1}$$

2.2.3. Identifying the core

A reduct is a subset of attributes RED $\subseteq P$, such that:

- $[x]_{\mathrm{RED}} = [x]_P$, that is the equivalence classes induced by the reduced attribute set RED are the same as the equivalence class structure induced by the full attribute set P.
- The attribute set RED is minimal, in the sense that $[x]_{(\mathrm{RED}-\{a\})} \neq [x]_P$ for any attribute $a \in$ RED, in other words, no attribute can be removed from set RED without changing the equivalence classes $[x]_P$.

A reduct can be considered as a sufficient set of features which mean to represent the category structure. The set of attributes which is common to all reducts is called the core. The core is the set of attributes which is possessed by every legitimate reduct and therefore consists of attributes which cannot be removed from the information system without causing collapse of the equivalence-class structure.

2.2.4. Calculating the value of attribute importance

Information is contained in the uncertainty. The stronger the uncertainty is, the more information is contained. In the subject of information science, the parameter, information entropy, is used to measure the degree of uncertainty and the uncertainty is expressed as the random in probability. If X is a set of infinite random variable P_i $(i = 1, \ldots, n)$, and the entropy of X is defined as:

$$H(X) = \sum_{i=1}^{n} P_i \log_2 \frac{1}{P_i} = -\sum_{i=1}^{n} P_i \log_2 P_i \tag{2}$$

Assuming two random variables X and Y, their joint and marginal probability distribution are $P(X, Y) = P\{X = x, \ Y = y\}$ and $P(X) = P\{X = x\}$, $P(Y) = P\{Y = y\}$, respectively. Under the condition of Y known, the conditional entropy of X is defined as:

$$\begin{aligned} H(X|Y = y) &= \sum_{i=1}^{n} p(y_i) H(X|Y = y_i) \\ &= -\sum_{i=1}^{n} p(y_i) \sum_{i=1}^{n} p(x_i|y_i) \log p(x_i|y_i) \end{aligned} \tag{3}$$

The above equation can be used to calculate the attribute importance value. This value reflects the how importance the specific attribute for the information system itself is. Therefore, all attributes can be ordered by their values, thus, the best attributes which can fully characterize the information system can be selected.

3. Risk evaluation mine ventilation system model

3.1. Support Vector Machine

Kernel-based techniques represent a major development in machine learning algorithms. Support Vector Machines (SVM) are a group of supervised learning methods that can be applied to classification or regression. The SVM algorithm is based on the statistical learning theory and represents an extension to nonlinear models of the generalized portrait algorithm developed by Vapnik (1998).

The core idea of SVM classification is to project 2-dimensional data into a constructed N-dimensional hyperplane, and then to determine an optimal separating hyperplane to separates the data into two or more categories. Fig. 1 shows the schematic for SVM classification.

Therefore, the key step in applying the SVM is to find a proper nonlinear mapping function (Kernel function) to finish the projection.

3.2. Establishing risk evaluation model

Essentially, the output of the risk evaluation model is a classification. A number of input parameters are inputted into the model. After data analysis, different corresponding risk ranks are determined to output. This finishes the evaluation or the prediction task. The procedure is shown in the following sections:

3.2.1. Nonlinear classification and kernel selection

The original optimal hyperlane algorithm proposed by Vapnik was a linear classifier. Boser et al. (1992) suggested a way to create nonlinear classifier by applying the kernel function. The feature of

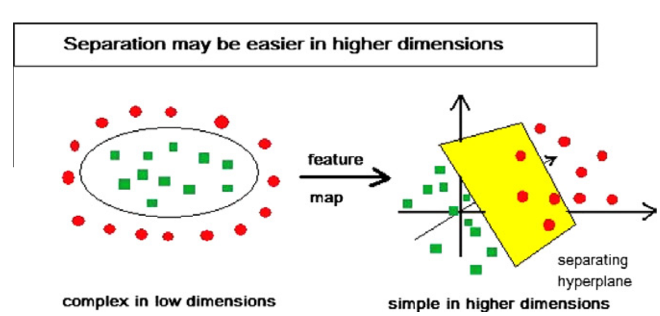

Fig. 1. Schematic diagram for SVM classification.