PAPER • OPEN ACCESS

Evaluation research of small and medium-sized enterprise informatization on big data

To cite this article: Na Yang 2017 *IOP Conf. Ser.: Mater. Sci. Eng.* **231** 012014

View the article online for updates and enhancements.

Evaluation research of small and medium-sized enterprise informatization on big data

Na Yang

Tianjin College, University of Science and Technology Beijing, Tianjin 301830,

China,yangna0534@163.com

Abstract. Under the background of big data, key construction of small and medium-sized enterprise informationization level was needed, but information construction cost was large, while information cost of inputs can bring benefit to small and medium-sized enterprises. This paper established small and medium-sized enterprise informatization evaluation system from hardware and software security level, information organization level, information technology application and the profit level, and information ability level. The rough set theory was used to brief indexes, and then carry out evaluation by support vector machine (SVM) model. At last, examples were used to verify the theory in order to prove the effectiveness of the method.

1 Introduction

Under the background of big data, small and medium-sized enterprise informatization level building needs to be focused on, but information construction cost will be high and whether the information cost inputs can bring benefit to small and medium-sized enterprises is still unknown. So how to effectively evaluate small and medium-sized enterprise informatization level is the research content in this paper[1].

SVM (Support Vector Machine, SVM) is a relatively new research results of artificial intelligence algorithms and has been used more and more widely. Its advantage is to overcome BP neural network fitting and easily into the local minimum problem when dealing with small sample data[2]. However, when SVM input data,the data redundancy cannot be determined, which increase the complexity of computing run. But rough set can effectively eliminate the noise in the input data [3]. So from this perspective, SVM algorithm and rough set theory are complementary inputting eliminating redundant indicators data to the SVM to train, and then predicting or evaluating test sample, which can reduce support in eliminating redundant information study and prediction burden, at the same time, improve the accuracy of the support vector [4]. Therefore, in this paper, the application of rough set and support vector machine (RS - SVM) are adopted to research on small and medium-sized enterprise informatization.

2. Rough Set - Support Vector Machine (SVM) Model Building

2.1 Model Building Description

In this paper, rough sets and support vector machine combined evaluation model is mainly used in the small and medium-sized enterprise informatization level evaluation, and the basic idea include application of theory of rough sets, eliminating redundant to effectively reduce noise of data, dividing the precise core attribute index data into the training sample and forecast sample and sending them to support vector machine (SVM) to learn and predict, and then evaluate.

2.2 Rough Set Theory and Attribute Reduction
Rough set theory is proposed by a polish mathematician Z. Pawlak in 1982, a new type of mathematical method dealing with vagueness and inaccuracy. When dealing with uncertain information, rough set does not need priority knowledge condition constraints, so it has been adopted in many fields.

In knowledge expression system $S = \langle U, C, D, V, f \rangle$, the reduction is different to decision tables $T = (U, A = C \cup D)$ It is not for the whole attribute set A, and only condition attribute set is reduced. Reduction of decision table is divided into two parts:

The first step: reduction for attribute set: if $P \subseteq C$, P is independent of attribute set D, and $Pos_p(D) = Pos_c(D)$, then P is the D reduction of C. All the intersection $RED_c(D)$ of D reduction in C is called the core of C "$Core(C)$"

The second step: reduction for attribute value: Given $T = (U, C \cup D)$ is consistent decision table, $P \subseteq C$ is the reduction D of C. Value reduction targets at reduction P, or we can say that attribute value reduction is related with each decision rule on decision table. Any condition attribute value of rules in decision table can be reduced and after that it can be consistent with rules, that means no inconsistency occurs.

2.3 SVM theory
SVM is a new machine learning algorithm, on the basis of statistical learning theory, proposed by Vapnik in 1995. There is a data set: $(x_1, y_1,), \cdots, (x_i, y_i) x_i \in R^n$ $y_i \in \{1. - 1\} i = 1,2, \cdots t,$ This data set can be distinguished by a hyper plane: $(\omega * x) + b = 0$ $x \in R^n, \omega \in R^n, b \in R$, at the same time it should be the best hyper plane. A quadratic programming problem is introduced in the process of seeking for hyper plane and the equation is as below:

$$min\varphi(\omega\omega = \frac{1}{2}(\omega * \omega)$$
$$s.t. y_i[(\omega * x_i) + b] \geq 1, i = 1,2, \cdots, t \qquad (1)$$

arrange optimize method was used to to solve the problems of the formula (1), and here is introduced into its dual problem, using nonlinear transformation(X) to tansfer original data set to the high dimensional space for data processing, and in this space for an optimal classification of the dual problem:

$$f(x) = sgn\left\{(\omega * X + b) = sgn\left\{\sum_{i=1}^{t} \partial_i y_i (X_i * X) + b\right\}\right\} \qquad (2)$$

2.4 Rough set - support vector machine (SVM) model for small and medium-sized enterprise informatization level evaluation
Rough set can preprocess the index involved in small and medium-sized enterprise information system and eliminate unimportant attributes index from informatization horizontal correlation, which reduces the time complexity of support vector machine (SVM), accelerate Support vector machine (SVM) learning process. To evaluate the small and medium-sized enterprise informatization level through rough sets and support vector machine (SVM) model, first rough set makes attribute reduction for evaluation index system in order to reduce unnecessary workload for evaluation index weight, then using analytic hierarchy process and entropy value method to calculate index weight, and finally input comprehensive informatization evaluation scores into the support vector machine to forecast and evaluate.

This paper first analyzes and researches on the influencing factors of small and medium-sized enterprise informationization level, collecting all kind of source data; next in the collected data, using rough sets to choose the data variables. At this stage, the obtained   sample set removes all the unnecessary variables and only retain the variables impact prediction accuracy; Then, using the analytic hierarchy process and entropy value method to calculate index weight and get comprehensive informatization evaluation scores which transfer to the support vector machine prediction system to

study and finally output the prediction results. After the model being built, in order to test whether it really can improve the prediction effect, this article uses the model to make empirical analysis.

3. Empirical Analysis
When filtering the evaluation index, this paper refers to the achievements of national informatization evaluation center and numerous research scholars, making comprehensive on the basis, and doing research on manufacturing enterprises from four aspects of informatization level of hardware and software security, information organization level, information technology application ability and profitability level.

This paper refers to enterprise informatization level evaluation index system of country small and medium-sized in reference [5-6] the system is as below table 1.

Table 1: informatization level evaluation index system of small and medium-sized enterprise

The first grade indicator	The second grade indicator	The third grade indicator
Hardware and software security level	Hardware and software level	Fixed assets level
		current assets level
		Computer level
		Database construction level
	Network level	Network performance level
		Network coverage
		Website construction level
	Security level	The level of safety investment proportion
		Security system construction level
	Labor security level	Employees' education level
		The informationization skill mastery level
		Electronical studying level
Information organization level	Status of informatization level	High-level attention
		National support level
		Informatization input level
		Informatization plan level
	institutions level	Setting level of information functions
		Coverage level of business process
Information technology application and profitability level	Marketing forecast level	Marketing research forecast level
		Designing and manufacturing level

		EPR level
	Informatization management level	Supply chain management
		Customer relationship management level
		OA level
	Marketing and information level	Marketing information level
		Decision-making information level
	profitability	Inventory proportion level
		Capital turnover levels
		Operating speed level
		Benefit growth levels
The informationization ability level	Information control ability level	Informatization plan ability level
		Informationization leadership level
		Informationization organization level
		Informationization control level
		Information technology innovation ability
	Information management module	Design and production capacity level
		Sales ability level
		Financial ability level
		Culture organizational ability level
		Technical system ability level

Table 1 shows that 40 specific indicators are included in this index system, such as information control ability level containing informationization plan ability level, informationization leadership level, information organization, information control, and information technology innovation ability and so on five indicators. Based on the original index system involved in table 1, we research the city's various types of 90 companies and obtain the firsthand data. Take information control ability level of indicators for example, as condition attributes, 5 sub indexes will obtain the original data processing, and according to the rough set theory,   divide them into four different levels, {4,3,2,1 }, grade respectively mean for {excellent, good, acceptable and bad}. The evaluation criterion of the enterprise is also divided into four types, as decision attribute of rough set D =,3,2,1 {4}, grade respectively mean for {good, better, general, poor}. The rest 11 have the same processing way.

4 Conclusion

This paper designed a kind of evaluation optimization strategy of rough set - support vector machines (SVM), overcoming the local minimum defects of the traditional BP network, and greatly improving the training speed. The enterprise informatization evaluation model was established based on rough set - support vector machine (SVM). Evaluating by this model was more accurate and more convenient than BP neural network model, and ruled out the subjective randomness of selected factors to give weight value, making the evaluation results more objective, so under the background of big data, the research method in this paper has important significance in theory and practical application.

References
[1]   BIG DATA INDUSTRY SUMMIT, http://www.idc quan.com/Special/bigdata2016/wenzhang2. html, (2016).
[2]   Li Guojie, Cheng Xueqi. "Big Data Research : Strategy&Policy Decision Research-Research Status and Scientific Thinking of Big Data", Bulletin of Chinese Academy of Sciences , 27(6), pp. 647-657, (2012).
[3]   Ma Lichuan, Pei Qingqi, Leng Hao. "Survey of Security Issues in Big Data". Radio Communications Technology, 41(1), pp.01-07, (2015).
[4]   Qian H, Mao Y, Xiang W, et al. "Recognition of human activities using SVM multi-class classifier". Pattern Recognition Letters, 31(2), pp.100-111, (2010).
[5]   Shalev-Shwartz S, Singer Y, Srebro N, et al. "Pegasos: primal estimated sub-gradient solver for SVM", Mathematical Programming, 127(1), pp.3-30, (2011).
[6]   Xie Hong,Cheng Haozhong, Niu Dongxiao. "Discretization of Continuous Attributes in Rough Set Theory Based on Information Entropy", Chinese Journal of Computers, 28(9), pp.1570-1574, (2005).