Alexandria Engineering Journal (2021) 60, 1519–1524

HOSTED BY

Alexandria University

Alexandria Engineering Journal

www.elsevier.com/locate/aej
www.sciencedirect.com

An evaluation index system for intellectual capital evaluation based on machine learning

Bei Yuan [a,b,*], **Hao Xia** [b], **Caicai Guo** [c]

[a] *Harbin Bank Postdoctoral Research Station, Harbin 150010, China*
[b] *Postdoctoral Station of Harbin Institute of Technology, Harbin 150001, China*
[c] *Northeast Agricultural University, Harbin 150038, China*

Received 17 August 2020; revised 20 October 2020; accepted 5 November 2020
Available online 8 December 2020

KEYWORDS

Intellectual capital;
Machine learning (ML);
Random forest (RF);
Support vector machine (SVM)

Abstract Currently, there is not yet a mature evaluation index system of intellectual capital among enterprises. The lack of such a system hinders the smooth transform of capital to enterprise value. Therefore, this paper attempts to set up an effective and objective evaluation index system for intellectual capital. First, the data on intellectual capital were collected from some enterprises from the Growth Enterprise Market (GEM). Next, the original data were preprocessed into 1770 effective pieces of data. On this basis, 13 indices were selected from three dimensions (e.g. human capital, structural capital, and relationship capital) of intellectual capital, forming an evaluation index system. After that, the evaluation index system was verified with two machine learning (ML) algorithms, namely, random forest (RF), and support vector machine (SVM). The results show that our evaluation index system can optimize the intellectual capital classification of enterprises, avoiding the subjective defects in qualitative evaluation. The research results shed important new light on the decision-making and scientific management of enterprises.

1. Introduction

Intellectual capital, such as patents, enables knowledge-based enterprises to remain competitive and encourages them to pursue innovation, thereby promoting the value creation of such enterprises [1]. Besides, intellectual capital is the basis for enterprises to formulate effective strategies against emergencies. For example, relational capital and structural capital, two important dimensions of intellectual capital, can be utilized to mitigate the risks arising from the unpredictable changes of the environment, and to respond to the complex and stochastic business affairs. Moreover, the competition between enterprises is essentially the competition of talents. To remain competitive, an enterprise must fully leverage human capital, an active driver of intellectual capital, and effectively combine capital, management, and innovation.

The pursuit of competitiveness and value creation has expanded the influence of intellectual capital from knowledge-based enterprises to all kinds of enterprises. The evaluation of

* Corresponding author at: Harbin Bank Postdoctoral Research Station, Harbin 150010, China; Postdoctoral Station of Harbin Institute of Technology, Harbin 15001, China.
E-mail addresses: yuanbei@hrbb.com (B. Yuan), xiah@hit.edu.cn (H. Xia), guocaicai1991@gmail.com (C. Guo).
Peer review under responsibility of Faculty of Engineering, Alexandria University.

https://doi.org/10.1016/j.aej.2020.11.006
1110-0168

intellectual capital can help companies increase their value [2]. Therefore, it is very meaningful for enterprises to measure their intellectual capital in an accurate and objective manner. With the development of financial science and technology, machine learning (ML), continues to change our lives and can provide more scientific calculation methods to support enterprises in assessing intellectual capital, predicting risks, and maximizing value [3,5]. However, the evaluation model must be coupled with a reasonable evaluation index system before making a scientific assessment of intellectual capital [6,7].

The intellectual capital of an enterprise is often intangible assets, rather than profitable products. Therefore, the existing evaluation index systems are unable to reflect the commercial value of intellectual capital. At present, most evaluation index systems for intellectual capital are purely theoretical, without sufficient data support. The indices are often extremely hard to collect. It is unsurprising that these systems are not highly practical.

In light of the above, this paper attempts to build a scientific evaluation index system of intellectual capital based on the relevant information, and verify its compatibility with the ML evaluation model. Firstly, the technical tools, e.g. the ML and random forest (RF) [8,9], of our research were introduced in details; Next, an evaluation index system was established for intellectual capital; After the research data were preprocessed, the proposed system was coupled with support vector machine (SVM) [4,10] to evaluate intellectual capital. Compared with existing research, the significance of this research is to prove that machine learning classification algorithms can be applied to the research field of intellectual capital indicators. Random forest model and support vector machine (SVM) model, as a classification algorithm in the machine field, can be applied to the intellectual capital evaluation system to obtain more accurate and reliable calculation results than traditional index classification, providing future research in the field of intellectual capital indicators New ideas. The research results provide new insights into the decision-making and scientific management of enterprises.

2. Methodology

The ML requires no prior knowledge of business rules or logic to evaluate intellectual capital. Instead, the business rules and logic are learned by an ML algorithm from a huge training set, and used to build an evaluation model for intellectual capital.

In the ML, the evaluation of intellectual capital is essentially a classification problem. During the evaluation, the intellectual capital is judged as good or bad. Good intellectual capital belongs to credible enterprises (positive examples), and bad intellectual capital belongs to untrustworthy enterprises (negative examples) [11]. To solve the classification problem, the relationship between the basic features of intellectual capital and enterprises and the good or bad of intellectual capital must be determined, and imported to the evaluation model.

On this basis, it takes four steps to solve a classification problem through supervised learning, namely, design Evaluation Index System, data acquisition, data preprocessing, and use RF classification and SVM evaluation for comparative processing. In this article, this question is defined as ML-based intellectual capital assessment. 1770 pieces of effective data about intellectual capital were mined through ML; the collected data was analyzed, cleaned and converted to make the evaluation more accurate; the training set was constructed through cross-validation, which is one of the three popular evaluation methods; Finally, select the appropriate learning algorithm to train the evaluation model and adjust the parameters.

2.1. Design of evaluation index system

As shown in Table 1, our evaluation index system for intellectual capital was constructed from three dimensions: human capital, structural capital and relationship capital. From the perspective of authority and utilization rate, this study selected the classic intellectual capital ternary theory in the classification of intellectual capital indicators. Guthrie proposed that the content of intellectual capital includes 18 elements in three categories, including organizational capital, customer capital, and human capital. Specifically, organizational capital is set to six elements

Table 1 The evaluation index system of intellectual capital.

Primary indices	Secondary indices	**Field description**
Human capital	Age structure	0 point if the age structure is bad; 1 point if the age structure is good; 2 points if the age structure is neutral
	Staff training	0 point if staff training is bad; 1 point if staff training is good
	Proportion of professional employees	Number
	Proportion of management talents	Number
Structural capital	Organizational structure	0 point if the organizational structure is bad; 1 point if the organizational structure is good; 2 points if the organizational structure is neutral
	Corporate culture	0 point if the corporate culture is bad; 1 point if the corporate culture is good; 2 points if the corporate culture is neutral
	Business efficiency	0 point if the business efficiency is bad; 1 point if the business efficiency is good; 2 points if the business efficiency is neutral
	Technological innovation	Number
Relationship capital	Corporate image	0 point if the corporate image is bad; 1 point if the corporate image is good; 2 points if the corporate image is neutral
	Investor relationship	Number of strategic alliances
	Consumer relationship	Consumer growth rate
	Supplier relationship	0 point if the supplier relationship is bad; 1 point if the supplier relationship is good; 2 points if the supplier relationship is neutral is 2
	Market share	Figure

of intellectual property, management process, corporate culture, financial relations, information and network systems, and management philosophy. Set customer capital as the seven elements of brand, customers, reputation, distribution channels, business cooperation, franchise contracts, and customer satisfaction; Set human capital as the five elements of education, training, work-related knowledge, and enterprising spirit.

2.1.1. Human capital

Human capital is affected by the external environment and many other factors, and thus difficult to quantify [12,13]. Here, the components of human capital are scrutinized to facilitate the evaluation of human capital. Besides age structure and staff training, two special indices were selected to measure human capital: the proportion of professional employees, and the proportion of management talents.

The former index stands for employees with professional technical ability and innovation ability as a proportion of all employees of the enterprise. The latter index stands for board members and management personnel as a proportion of all employees of the enterprise. The specific number of the two indices are available in the annual reports of listed enterprises.

2.1.2. Structural capital

Structural capital refers to the integration and cooperation mechanism of the enterprise. Under this mechanism, individuals are encouraged to exchange professional knowledge and experience, forming the collective wealth of the enterprise. Structural capital links up human capital with relationship capital, and promotes the mutual transform between various elements. Here, structural capital is measured by four indices, including organizational structure, corporate culture, business efficiency, and technological innovation.

The organizational structure stands for the basic skeleton of enterprise management, involving all departments and positions in the enterprise. It plays an important role in organizing and coordinating the cooperation of all parties in the enterprise.

The corporate culture, as the key to the creation of enterprise values, is the sum of the values, behavior norms and thinking patterns that are typical of the enterprise. The corporate culture is formed and followed by the enterprise through long-term practice [14].

The business efficiency reflects how well the enterprise responds to the fierce competition in the market. The efficiency level depends on the overall planning and strategy of the enterprise for long-term development.

The technological innovation is mainly demonstrated by the research and development (R&D) efforts of the enterprise. In general, a high-value patent requires high human, material and financial inputs into the R&D. Since the data on the latter two inputs are not available, this paper decides to measure technological innovation by the R&D results, including intellectual properties and core technologies.

2.1.3. Relationship capital

Relationship capital refers to the capital invested by the enterprise to achieve its business goals, and to forge and maintain relationship with stakeholders. To survive the fierce competition, the enterprise must cultivate a good corporate image, occupy a large market share, and maintain the relationship with investors, consumers, and suppliers.

The corporate image is the impression of consumers and other stakeholders on the services and products provided by the enterprise, i.e. the thing that differentiates the enterprise from other enterprises. The corporate image encompasses such elements as corporate reputation, brand awareness, social responsibility, etc.

The investor relationship stands for the relationship between the enterprise and its investors. A good investor relationship improves the participation of investors and protects their interests. Here, this relationship is characterized by the number of strategic alliances between the enterprise and its business partners.

The consumer relationship reflects the loyalty, satisfaction, and growth rate of consumers. It has a nonnegligible impact on enterprise performance. The main consumers are often disclosed in the annual reports of enterprises. Here, this relationship is characterized by the growth rate of consumers.

The supplier relationship stands for the business contacts and cooperative ties of the enterprise. A good supplier relationship is the prerequisite for the normal operation of the enterprise, especially in the manufacturing field.

The market share stands for the proportion of the market occupied by the enterprise. To gain a large market share, the enterprise must set up and maintain commercial channels, lay down proper marketing strategies, and select a suitable selling method [15].

To verify its rationality, the above evaluation index system was applied to evaluate the intellectual capital of some enterprises from the Growth Enterprise Market (GEM), the second board of the China stock market. Most of them belong to emerging industries like information technology (IT), new materials, new energy, and so on. Intellectual capital plays an important role in these industries. As a result, the GEM enterprises tend to disclose their intellectual capital in annual reports.

The selected enterprises were listed on the GEM for three consecutive years from 2015 to 2019, according to a list of GEM enterprises published by Shenzhen Stock Exchange. The data on intellectual capital were collected from the annual reports and websites of the selected enterprises. Due to the massive amount of intellectual capital information, the author collects it through web crawler technology, sets "company name + keywords" to start data capture, removes duplicate information, and uses "intellectual capital" and the keywords in Table 1 to control the information collection Direction to form a network dataset. According to the above classification standards, the annual report information is sorted and sorted to form an annual report data set. Finally, a total of 1770 pieces of effective data were generated. The basic information of effective samples is shown in Table 2.

Based on the evaluation index system, the 1770 pieces of effective data on intellectual capital were classified empirically. For the fairness of the classification, the selected categories include various fields, such as new materials, biomedicine, and equipment manufacturing. The workflow of the empirical classification is shown in Fig. 1.

2.2. Data preprocessing

The original data on the 13 aspects (secondary indices) of intellectual capital are rather messy, and varied in dimensionality and form. To make the data correct, accurate and integral,

Table 2 The basic information of effective samples.

Type of enterprise	Proportion%
IT	11.65
New materials	12.54
New energy	13.65
Aerospace	18.12
Biomedicine	19.43
Equipment manufacturing	17.69
Others	6.92

the 1770 pieces of data were preprocessed through analysis, cleaning, and transform. There is no strict order between the specific steps of the preprocessing. To improve data quality, some steps were executed multiple times.

For convenience, some variables that are too complex were combined properly. A total of 47 missing items were identified and supplemented, providing a guarantee for accurate prediction. In addition, the similarity between samples was quantified in quadratic form. On this basis, the input variables that are neither zero or one were normalized by the preprocessing module on the scikit-learn platform. Through the preprocessing, the research data became more suitable for model learning [16].

2.3. RF classification

Proposed by Leo Breiman, the RF selects n random samples through bootstrap resampling from the original training set M, generates an ensemble of n decision trees from the self-service sample sets, and classify new data based on the scores voted by the decision trees [17,18]. The RF can be implemented in the following steps:

Step 1. From the original training set M, n new self-service sample sets are extracted randomly by self-service method, and used to construct K decision trees. The unextracted samples each time are referred to as out-of-bag data.

Step 2. From each node on each tree, a variable B is randomly selected. The variable with the strongest ability to classify the variables is identified against a preset threshold H.

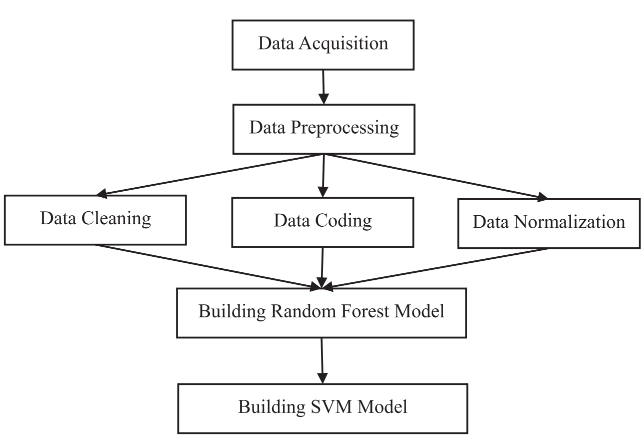

Fig. 1 The workflow of intellectual capital evaluation.

Step 3. Each tree grows to the maximum without being pruned. Then, multiple decision trees are generated to form an ensemble. Then, the new data are differentiated by a RF classifier. The classification result depends on the number of votes of the classifier.

In this research, the 13 secondary indies in Table 1 are taken as the initial independent variables, and the level of intellectual capital is considered as the dependent variable (1 point if the intellectual capital is good, and 0 point if the intellectual capital is bad).

Then, the 1770 pieces of data were divided into a training set and a test set at the ratio of 3:2. That is, the training set contains 1062 pieces of data, while the test set involves 708 pieces of data. The pseudocode of the RF and the classification accuracy are as follows:

```
> plot(model)
> pred = predict(model,testdata[,-13])#forecast test data
> accuracy = sum(pred = =testdata$ yes no)/nrow(testdata)#c
alculation accuracy
> accuracy
[1] 0.687
```

The importance of each index obtained by the RF is shown in Table 3.

The modelling results show that 68.7% of the 13 indices were classified accurately, which is highly satisfactory. Note that the importance of every index was positive. The proportion of professional employees, technological innovation, and consumer relationship had relatively high importance. Hence, the three indices are key metrics of intellectual capital. By contrast, age structure and staff training had relatively low importance, and are not highly representative of intellectual capital.

2.4. SVM evaluation

Despite its accuracy and stability, the ensemble classifier of the RF faces two defects: First, the RF is prone to overfitting when the decision trees try to fit the dataset; Second, the RF is not

Table 3 The importance of each index in the evaluation index system.

Serial number	Evaluation indices	Importance
1	Age structure	1.21910764
2	Staff training	2.30726431
3	Proportion of professional employees	68.32792326
4	Proportion of management talents	24.8978824
5	Organizational structure	24.63778757
6	Corporate culture	9.03028872
7	Business efficiency	25.78855386
8	Technological innovation	44.29527989
9	Corporate image	10.86224907
10	Investor relationship	14.13486425
11	Consumer relationship	40.54136437
12	Supplier relationship	32.22180532
13	Market share	7.60534417

good at dealing with unbalanced data [19]. To solve the defects, the SVM was introduced to evaluate the intellectual capital again based on our evaluation index system, because the SVM can reduce the possibility of overfitting and outshine the RF in predicting unbalanced data [20].

To optimize the generalization ability, the SVM strikes a balance between complexity and learning ability, and relies on optimization to solve ML problems, thereby overcoming the curse of dimensionality [21]. The algorithm enjoys great advantages in handling small-sample, nonlinear and high-dimensional problems.

In this research, the 1770 pieces of data were randomly divided into a training set and a test set. After the division, the training set contains 1239 (70%) pieces of data, while the test set involves 531 (30%) pieces of data. Meanwhile, the 13 secondary indies in Table 1 were still taken as the initial independent variables, and the level of intellectual capital as the dependent variable. The pseudocode of the SVM and the prediction accuracy are as follows:

```
> plot(model)
> pred = predict(model,testdata[,-13])#forecast test data
> accuracy = sum(pred = =testdata$ yes no)/nrow(testdata)#c
alculation accuracy
> accuracy
[1] 0.834
```

It can be seen that the SVM correctly predicted 92% of the 531 pieces of test data, which higher than the accuracy of the RF (83.4% > 68.7%). This means the SVM classifier can effectively classify the indices of intellectual capital, and verifies the feasibility of our evaluation index system.

3. Conclusions

This paper sets up an evaluation index system of intellectual capital, and verifies its effectiveness with two separable ML algorithms, namely, RF and SVM. Both algorithms take the 13 secondary indices as independent variables and the level of intellectual capital disclosed by enterprises as the dependent variable. The results show that 10 of the 13 (68.7%) secondary indices were classified accurately by the RF and SVM. Therefore, the proposed evaluation index system is suitable for assessing the level of intellectual capital of the enterprise.

Based on the research results, two suggestions were put forward to improve intellectual capital:

(1) The government should set up a communication platform for enterprises to share and exchange experience and knowledge, laying the basis for open innovation.
(2) The enterprise should no longer focus solely on human capital. Equal attention should be paid to structural capital and relationship capital. For example, the enterprise needs to attach importance to teamwork, mutual trust and other aspects of the structural capital.

The main goal of this research is to build a classification system of intellectual capital evaluation indicators, which has certain limitations. In the classification, the intellectual capital is set as human capital, relational capital, and structural capital, and only some specific indicators that affect the value of the enterprise are considered when setting the subdivision indicators under the general indicators. In future research, we can appropriately refine the indicators, introduce other dimensional variables, and further explore the mechanism of corporate intellectual capital by constructing models.

Declaration of Competing Interest

The authors declared that there is no conflict of interest.

Acknowledgement

This paper is made possible thanks to the generous support from the General Project of National Natural Science Foundation of China (Grant No.: 71532004; 71672046; 71490724).

References

[1] R.D. Wimmer, J.R. Dominick, Mass Media Research: An Introduction, Wadsworth, Belmont, California, 1994, pp. 165–178.
[2] L.A.M. Chowdhury, T. Rana, M. Akter, M. Hoque, Impact of intellectual capital on financial performance: evidence from the Bangladeshi textile sector, J. Accounting Organizational Change 14 (11) (2018) 429–454.
[3] D.B. Tran, D.H. Vo, Should bankers be concerned with Intellectual capital? A study of the Thai banking sector, J. Intellectual Capital 19 (5) (2018) 897–914.
[4] K. Gorur, M.R. Bozkurt, M.S. Bascil, F. Temurtas, GKP signal processing using deep CNN and SVM for tongue-machine interface, Traitement du Signal 36 (4) (2019) 319–329.
[5] I. Chouidira, D.E. Khodja, S. Chakroune, Induction machine faults detection and localization by neural networks methods, Revue d'Intelligence Artificielle 33 (6) (2019) 427–434.
[6] I. Chouidira, D.E. Khodja, S. Chakroune, Continuous wavelet technique for detection of broken bar faults in induction machine, Traitement du Signal 36 (2) (2019) 171–176.
[7] Y.R. Yang, Do aggressive pro forma earnings-reporting firms have difficulty disclosing intellectual capital? Australian evidence, J. Intellectual Capital 19 (5) (2018) 875–896.
[8] S. Murugan, G. Kulanthaivel, V. Ulagamuthalvi, Selection of test case features using fuzzy entropy measure and random forest, Ingénierie des Systèmes d'Information 24 (3) (2019) 261–268.
[9] S.R. Khonde, U. Venugopal, Hybrid architecture for distributed intrusion detection system, Ingenierie Systemes d'Information 24 (1) (2019) 19–28.
[10] S.P. Deore, A. Pravin, Histogram of oriented gradients based off-line handwritten Devanagari characters recognition using SVM, K-NN and NN classifiers, Revue d'Intelligence Artificielle 33 (6) (2019) 441–446.
[11] G. Ginesti, A. Caldarelli, A. Zampella, Exploring the impact of intellectual capital on company reputation and performance, J. Intellectual Capital 19 (5) (2018) 915–934.
[12] M.A.I. Cisneros, F. Hernandez-Perlines, Intellectual capital and Organization performance in the manufacturing sector of Mexico, Manag. Decis. 56 (8) (2018) 1818–1834.
[13] K. Krippendorff, Reliability in content analysis: Some common misconceptions and recommendations, Human communication research 30 (3) (2004) 411–433.
[14] E. Pedro, J. Leitão, H. Alves, Back to the future of intellectual capital research: a systematic literature review, Manag. Decis. 56 (11) (2018) 2502–2583.

[15] M.A. Boujelbene, H. Affes, The impact of intellectual capital disclosure on cost of equity capital: A case of French firms, J. Economics Finance Administrative Sci. 18 (34) (2013) 45–53.
[16] L. Bottou, F.E. Curtis, J. Nocedal, Optimization methods for large-scale machine learning, SIAM Rev. 60 (2) (2018) 223–311.
[17] L. Breiman, J. Friedman, C.J. Stone, R.A. Olshen, Classification and regression trees, CRC Press, 1984.
[18] M. Brijain, R. Patel, M. Kushik, K. Rana, A survey on decision tree algorithm for classification, Int. J. Eng. Development Res. 2 (1) (2014) 13–19.
[19] S. García, S. Ramírez-Gallego, J. Luengo, J.M. Benítez, F. Herrera, Big data preprocessing: methods and prospects, Big Data Analytics 1 (1) (2016) 9.
[20] P. Gogas, T. Papadimitriou, A. Agrapetidou, Forecasting bank failures and stress testing: A machine learning approach, Int. J. Forecast. 34 (3) (2018) 440–455.
[21] Hong, H., Kubik, J.D., Scheinkman, J.A. (2012). Financial constraints on corporate goodness (No. w18476). National Bureau of Economic Research, 2011(1), 72-81.