Intellectual Capital of Enterprises in Thailand: Measurement Model by Baysean Network Algorithm

Anongnart Srivihok
Department of Computer Science, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand
email: fsciang@ku.ac.th

Abstract

Intellectual Capital (IC) is one type of capital which is known as an intangible asset of an organization. Therefore, IC is claimed to be a valuable asset in the organization. However, the study of IC was not widely conducted in Thailand. IC includes human capital, structural capital and customer capital. This paper applies data mining techniques, classification algorithms for generating IC model of organizations in Thailand. Three candidate classification algorithms including Decision tree (ID3), Decision Tree (C4.5) and Bayesian Network were compared for the prediction powers in this study. Data set was obtained form a survey of 216 organizations located in the central part of Thailand. Results show that Bayesian Network has the highest prediction power. The accuracy of this IC model is about 83% which is good. The implication of this model is also suggested.

Keyword: Intellectual Capital, data mining, decision tree, Bayesian Network

1. Introduction

In the last two centuries, it was common knowledge that there were only two factors of production which include labor and capital. At present, information and knowledge are recognized as the primary wealth-creating assets, just as the latter two replaced land and labor in last two centuries. Further, the development of new technologies in the 20th century has transformed the majority of wealth-creating work from physical-based to "knowledge-based" dimensions.

Technology, information and knowledge are now the key factors of production and services success. With the increased flow of information and the global work force, knowledge and expertise can be transported instantaneously around the world, and any advantage gained by a company can be reversed by competitive improvements overnight. The company will proceed to its process of innovation by combining market and technology know-how with the creative talents of knowledge workers. This can be used to win over a large stream of competition and its ability to generate value from knowledge. We are now in a knowledge society heading to a knowledge economy.

Knowledge is considered as an Intangible Asset. The concept of Intangible Asset is very broad. In the concept given by the most important accounting standards, it is assumed that an asset (tangible or intangible) must meet three requirements: (1) generates profits in the future, (2) controlled and managed by the company and (3) has a value or a cost that could be measured. The term intangible is related to those items without material background such as patents, software, database, policies, manual, and procedures [1].

The international standards have developed rules trying to reflect these intangible assets into the book-values under some sections: research investments, patents, trademarks, use rights, etc. But there are other items not being considered although they accomplish the three rules, for example, the company client's portfolio or the human resources. So it is necessary that a new categorization for the intangible assets is established.

The term Intellectual Capital was introduced by Leif Edvinsson [3] for measuring these intangible assets. However, the Intellectual Capital is not widely considered as an asset by many organizations because of the difficulty and the uncertainty of measurement of future profits.

Intellectual Capital is essential as well in order to compare different companies, to estimate their real value or even to control their improvement year-to-year. But to measure a concept is necessary to determine exactly what it is.

The interest on managing the intangible assets was derived in the development of different methods of measuring Intellectual Capital. One of the reasons is the existing need of the companies to improve the way they manage the things that generate value and give back some benefits.

2. Background of the study

In the era of a knowledge-based economy, the organizational competitiveness is based on the accumulation of knowledge and intellectual capital (IC) because organizational future assets tend to depend on IC instead of traditional physical assets. Therefore, they are willing to audit and measure IC and make it transparent to manage. IC measurement has been developed by many organizations and researchers, initiated by [7] and [6]. Firstly, Skandia, a Swedish financial services company, was the first company to substitute traditional financial data in its annual report with IC value [3]. According to the Skandia model, IC consists of two main elements of human and structural capital.

Human capital combines knowledge, professional and social competence, capabilities, expertise, creativity, motivation, and leadership of an organisation's staff. On the other hand, structure capital covers internal processes, infrastructure (e.g. information technology, management database) culture, patents, training programs, and organisational strategies that support its core competence [3].

Brooking [1] proposed IC as an integration asset of human-centred assets, infrastrutural assets (e.g. processes, methods, and technology), intellectual property assets (e.g. copyrights and patents) and market assets. Later, Roos and his colleagues [4] presented an IC model with the components of human capital (e.g. intellect, skill, creativity, work procedure), organisational capital (e.g. system, IP, processes, databases, values, and culture), and relational capital. Relational capitals are assets derived from good relationships with suppliers, customers, partners, networks, regulators and interrelating stakeholders. Among others, customer capital (e.g. customer relationships, loyalty) is accepted as the most vital asset [2].

Sveiby [8] proposed direct intellectual capital methods (DIC), market capitalization method (MCM), return on assets methods (RA) and scorecard methods (SC), number of times in training (days per year), and annual sales per customer.

In Thailand, the IC measurement of SME had been conducted by Srivihok [5]. A measurement model and a qualitative index system of IC were designed in order to provide a good tool for organisations to manage their IC. The study found that there is a significant relationship between the scores of the three IC elements (i.e. human capital, structural capital, and customer capital) and its business performance.

The past studies of Srivihok [5], intellectual capital (IC) model was adapted for use in this present study. The IC model (figure 1) consists of human, structural and relational capital. The indicators for human capital are people competence, competence improvement, staff structure, improvement of personal capacity and innovation, and stability. Structural capital is divided into process technology and IT penetration, business philosophy, organisation structure, and intellectual property. Relational capital consists of customer base, customer loyalty, market proximity and marketing effectiveness. Innovation capital is not included in this study because it is a subset of structural capital.

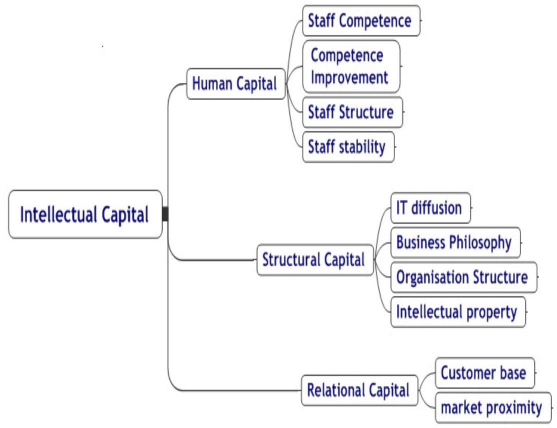

Figure 1. Intellectual Capital Measurement Model

3. An empirical analysis of Intellectual Capital measures

The research was designed to be an explanatory, cross sectional, statistical, mail survey. The survey study investigated in Thai organizations. Respondents of the survey were staff in the organisations, and the survey focused on measuring attributes of Intellectual Capital as shown in Table 1, more details on the questionnaire can be accessed from a web site: www.smexpert.kaetsart.org/ric. It was hypothesized that Intellectual Capital is related to human capital, structural capital and relational capital which would affect the organisation's performance. For this purpose, Intellectual Capital would be measurable by 24 factors which are the attributes of human, structural and relational capital.

Table 1. Attributes of Intellectual Capital used in Intellectual Capital measurement

IC Components	**Attributes**
Staff Competency	% staff with knowledge in working
	% staff using internet
	% experienced staff
Competency Improvement	Organizational support for a learning organization
	% staff get appropriate training
	% staff apply acquired knowledge in working
Staff Structure	% staff with long working years
	ability to replace a staff

IC Components	Attributes
	staff satisfaction to manager
Staff Stability	Good work environment
	% staff turnover
Production technology and IT diffusion	% computers / staff
	% IT investment
	Organization database
Business Philosophy	Investment in planning and implementation of the plan
	Organization defines the missions clearly
	Customer oriented organization
Organizational Structure	Organization structure (2, 3 or 4 levels)
Intellectual Property	Number of Intellectual Properties
	% research and development expenses / revenue
Customer base	% customer satisfaction
	% customer loss
Market Proximity	number of communication channels in organizations (internet, phone, fax, mail, direct)
	% number of communication channels in organizations

.

3.1 Data Collection

The study of Intellectual Capital measurement was focused in both public and private organisations in Thailand. There were two parts of the questionnaire: intellectual capital section and organisational background. The responses in the first section were measured on a 10-point semantic differential scale with 1= strongly disagree, and 10 = strongly agree. The first section contains questions for evaluation attributes. The background information section was designed to obtain information on organisation characteristics including type of industry, size, and years of services. The first part of the questionnaire was developed from the IC model proposed by Srivihok [5] as shown in Table 1. Data collection was conducted by mail survey to 800 public and private organisations. Names and addresses of private organisations were obtained from the list of Import-Export companies provided on the website of the Department of Extension, Ministry of Commerce, Thailand. The list of public organisations was obtained from government agencies. There were about 216 respondents from the survey which was about 27%. This rate was considered adequate for a mail survey.

For the organisational characteristics, majority of them were SME, about 42.9% were small enterprises with less than 50 staff, 23.8% were medium enterprises with 51-200 staff, and 20.5% and 12.8 % were large enterprises with 201-1000 staff and more than 1000 staff, respectively. The industry type and number of participation are shown in Table 2.

Table 2. shows that the majority of type of industry was manufacturing (32.60%), the next was education 14.8% and the smallest was tourism which was about 1.00%. For the organizational type, about 24.8% were government, 3.6% were state enterprises and 71.6% were private enterprises which were the majority of the population of this study.

Table 2. Organizations segmented by industry.

Industry	Percent
Agriculture	3.30%
Chemistry	10.10%
Education	14.80%
Energy	1.40%
Finance	2.90%
Food	3.30%
IT	3.80%
Manufacturing	32.60%
Services	8.60%
Tourists	1.00%
Textiles	1.40%
Others	16.80%

The Intellectual Capital (IC) scores of each organization were calculated by summing all attribute scores obtained from the survey as shown in Table 1. The calculation was done as follows:

$$IC = \sum_{k=1}^{n} a_k$$

IC = Intellectual Capital score
a_k = value of each attribute

Each organization was classified into class A, B or C by using IC scores. Table 3 presents the classification of Intellectual Capital score. There were 50 organizations in the low class, 133 organizations in the medium and 33 organizations in the high class.

Table 3 Classes of IC scores (n=216)

Class	IC scores
low	Less than 116
medium	165-171
high	more than 171

3.2 Data analysis

WEKA software Version 3.5.6 (Waikato Environment for Knowledge Analysis) [9] was used to analyze the data from the survey. Cross-validation of data set was conducted by using the Hold-Out Method. The data set was divided into two sub-sets: training set and testing set. The training set consisted of 80% of the total data while the remaining 20% was employed in the testing set. The three main classification algorithms used were Decision tree (ID3), Decision tree (J48), and Bayesian Network. There were 216 instances and 24 attributes were analyzed by these three algorithms. The IC models were developed on the basis of three candidate algorithms and were used to predict the class of organisations.

3.3 Data Mining Techniques

There were three data classification algorithms used in this study. They include Decision Tree (ID3), Decision Tree (C4.5) and Bayesian Network. There are reasons in using these algorithms include decision tree was claimed to be effective in generating a tree for decision making, it provides the sequences of predecessor and successor nodes. Also, Bayesian Network is an algorithm which based on the probability theory of Bayes. This network shows the relationship between the nodes (attributes) of the model. The fundamental of each algorithm is as follows:

3.3.1 ID3 Algorithm

ID3 (Iterative Dichotomiser 3) is an algorithm used to generate a decision tree. It is a greedy algorithm that grows the tree top-down. Also each node selecting the attribute that best classifies the training data. The algorithm is based on Occam's razor: it prefers smaller decision trees, and is therefore a heuristic. Occam's razor is formalized using the concept of information entropy.

The basic ideas of ID3 are that:

- In the decision tree, each node corresponds to a non-categorical attribute and each arc to a possible value of that attribute. A leaf of the tree specifies the expected value of the categorical attribute for the records described by the path from the root to that leaf.
- In the decision tree, each node should be associated with the non-categorical attribute which is most informative among the attributes not yet considered in the path from the root.
- Entropy is used to measure how informative a node is.

3.3.2 C4.5 Algorithm

C4.5 algorithm is Quinlan's extension of his own ID3 algorithm for generating a decision tree [10]. This algorithm recursively visits each decision node, selecting the optimal split, until no further splits are possible. The C4.5 algorithm is not restricted to binary splits, it produces a tree of a more variable shape. By default it produces a separate branch for each value of the categorical attribute.

C4.5 algorithm uses the concept of information gain or entropy reduction to select the optimal split. Main improvements included in C4.5 deal with the pruning methodology and the processing of numeric attributes.

3.3.3 Bayesian Network

Bayesian Network [10] or Bayes Net is a specific type of graphical model which is a directed acyclic graph. That is, all of the edges in the graph are directed and there are no cycles. A Bayesian Network can be used to compute the conditional probability of connected nodes, given values assigned to the other nodes. A Bayesian Network can be used as a classifier that gives the posterior probability distribution of the class node, given the values of other attributes.

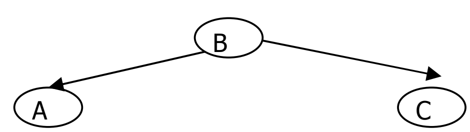

Figure 1 Example of Bayesian Network

Figure 1 presents a Bayesian Network model which is represented by a set of edges, e.g.., E={(B,A),(B,C)}. The edges in the Bayesian Network encode a particular factorization of the joint distribution. In this example, the joint distribution of all the variables, as factored by this Bayesian Network, is:

$$P(A,B,C) = P(A \mid B) \bullet P(B) \bullet P(C \mid B)$$

A Bayesian Network is a carrier of the conditional independencies of a set of variables, not of their causal connections. However, causal relations can be modeled by the closely related causal Bayesian Network.

3.4 Evaluations

The prediction performances of four algorithms are evaluated by using precision, recall, F-measure, accuracy and Root mean-squared error (RMSE). Precision and recall appropriateness have been used extensively to evaluate the retrieval performance of information retrieval algorithms. However, a more careful reflection reveals problems with these two measures. First, the proper estimation of maximum recall for a query requires detailed knowledge of all the documents in the collection. With large collections, such knowledge is unavailable which implies that recall can not be estimated precisely. Secondly, recall and precision are related measures which capture different aspects of the set of retrieved documents.

3.4.1 Precision

Precision is the measurement of how much of the data returned is correct.

$$\text{Precision}=\frac{\text{Number of correct answers given by system}}{\text{Number of answers given by system}}$$

3.4.2 Recall

Recall is the measurement of how much relevant data the system has.

$$\text{Recall} = \frac{\text{Number of correct answers given by system}}{\text{Total number of possible correct answers}}$$

3.4.3 F-measure

Precision and Recall stand in opposition to each other. As precision goes up, recall usually goes down. The F-measure combines the two values.

$$\text{F-measure} = \frac{(B^2+1)*(\Pr ecision * \operatorname{Re} call)}{B^2*(\Pr ecision + \operatorname{Re} call)}$$

When B = 1, precision and recall are weighted equally.

When B is > 1, precision is favored.

When B is < 1, recall is favored.

3.4.4 Root mean-squared error (RMSE)

$$\text{RMSE} = \sqrt{\frac{(a_1-c_1)^2+(a_2-c_2)^2+\ldots+(a_n-c_n)^2}{n}}$$

The mean-squared error is one of the most commonly used measures of success for numeric prediction. This value is computed by taking the average of the squared differences between each computed value (c_i) and its corresponding correct value (a_i). The root mean-squared error is simply the square root of the mean-squared-error. The root mean-squared error gives the error value the same dimension as the actual and predicted values. The smaller the values of RMSE, the better the power of prediction.

4. Results

The prediction powers of three candidate algorithms are revealed in Table 4, 5, 6, and 7 as follows.

Table 4 presents the results from predicting the Intellectual Capital Class of an organization using Bayesian Network. For organizations classified as Medium (score between 116-165), the prediction value was good, given the precision value of 0.89, the recall value of 0.84, and the F-Measure value of 0.86. These values are the highest among the three classes.

For organizations classified as High (score higher than 165), their prediction value was good, given the precision value of 0.82, the recall value of 0.82, and the F-Measure value of 0.82. For organizations classified as Low (score less than 116), their prediction value was also good, given the precision value of 0.69, the recall value of 0.84, and the F-Measure value of 0.76.

Table 4. Prediction Value of Bayesian Network

Class	Precision	Recall	F-Measure
High	0.82	0.82	0.82
Medium	0.89	0.84	0.86
Low	0.69	0.84	0.76

Table 5 presents the results from predicting the Intellectual Capital Class of organizations using Decision Tree algorithm (C4.5). For organizations classified as Medium (score between 116-165), the prediction value was good, given the precision value of 0.84, the recall value of 0.80, and the F-Measure value of 0.82. These values are the highest among the three classes.

For organizations classified as High (score higher than 165), their prediction value was good, given the precision value of 0.68, the recall value of 0.71, and the F-Measure value of 0.69. For organizations classified as Low (score less than 116), their prediction value was also good, given the precision value of 0.72, the recall value of 0.81, and the F-Measure value of 0.76.

Table 5. Prediction Value of Decision Tree (C4.5)

Class	Precision	Recall	F-Measure
High	0.68	0.71	0.69
Medium	0.84	0.80	0.82
Low	0.72	0.81	0.76

Table 6 presents the results from predicting the Intellectual Capital Class of organizations using Decision Tree algorithm (ID3). For organizations classified as Medium (score between 116-165), the prediction value was good, given the precision value of 0.84, the recall value of 0.84, and the F-Measure value of 0.84. These values are the highest among the three classes.

For organizations classified as High (score higher than 165), their prediction value was good, given the precision value of 0.76, the recall value of 0.70, and the F-Measure value of 0.73. For organizations classified as Low (score less than 116), their prediction value was also good, given the precision value of 0.62, the recall value of 0.76, and the F-Measure value of 0.74.

Table 6. Prediction value of Decision Tree (ID3)

Class	Precision	Recall	F-Measure
High	0.76	0.70	0.73
Medium	0.84	0.84	0.84
Low	0.72	0.76	0.74

From the comparison of the prediction powers of the three candidate algorithms, Table 7 shows that Bayesian Network has the highest power in accuracy (83.80%) and the lowest RMSE (0.29). The smaller the values of RMSE, the better the power of prediction is.

Table 7. Comparison of prediction accuracy by three classification algorithms

Measures	DT(ID3)	DT(C4.5)	Bayes Net
Accuracy	78.70 %	77.78 %	83.80 %
RMSE	0.36	0.38	0.29

Then Bayesian Network was used to generate the model of Intellectual Capital from the data set of 216 organizations. Then the sequence of the attributes of the model are shown in Figure 2.

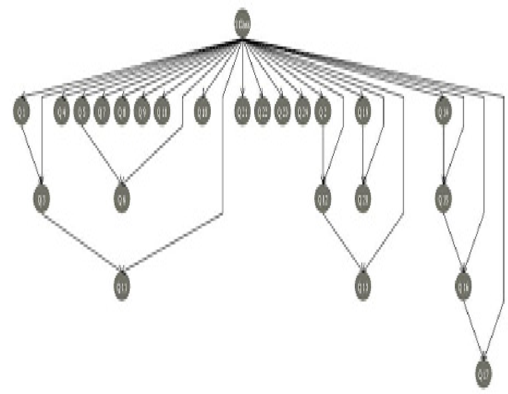

Figure 2. The Intellectual Capital model from Bayesian Network.

From the IC model in Figure 2, all attributes are related to Intellectual Capital directly and indirectly. The attributes which are directly related to IC include 1. staff with sufficient knowledge, 2. staff using the internet 3. Organizational support for a learning organization, 4. staff get appropriate training, 5. staff with long working years, 6. ability to replace a staff, 7. staff satisfaction on manager 8. Good working environment, 9. Organization Database for management 10. Organization structure, 11. number of Intellectual properties 12. customer satisfaction, 13. customer loss, 14. number of communication channels in organizations, and 15. investment on marketing/revenue. Attributes which indirectly influence IC include 1. experienced staff, 2. staff apply acquired knowledge in working, 3. staff turnover, 4. computers/staff, 5. IT investment, 6. research and development expenses/revenue, 7. Investment in planning and implementation, 8. Organization Mission, and 9. Customer relations.

As the IC model is divided into small parts. Figure 3 shows associations and sequences of attributes. Therefore, the Organization Database for management (Q14) dominated other attributes including Investment in planning and implementation (Q15), Organization Missions (Q16), and Customer relations (Q17), respectively.

Other relationships among attributes include (1) number of Intellectual properties associated to R&D investment, (2) staff get appropriate training relates to staff applying acquired knowledge in working**,** (3) staff with knowledge dominated in working and experienced staff dominated staff turnover.

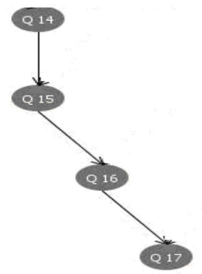

Figure 3. A part of IC model showing the relationships between IC attributes.

5. Conclusions

This research has compared the three candidate classification algorithms, *Decision tree (ID3), Decision Tree (C4.5) and Bayesian Network.* The highest perfomance, *Bayesian Network* applied the data mining technique to develop the IC Model. Employing organizational data from both private and public organizations, the IC model was shown to have good predictive power with good accuracy rate (83.8%).

The IC model can be applied in enhancing understanding of Intellectual Capital attribute patterns. It can also be used for developing and improving the intellectual capital to fit the needs of organizations. This may in turn lead to an increase in the intangible assets, and thus increase the intellectual properties of organizations in the near future.

Future research work might include improvement of the IC model by using or incorporating other data mining techniques such as Naïve Bayes Tree, Decision Forest or association rules in model development to make the model more effective.

6. Acknowledgement

This project was supported by the research grant in year 2007 provided by the Research and Development Institute of Kasetsart University. Also, we would like to thank Dr. Cappia for editing this paper.

7. References

[1] Brooking, A. "Intellectual capital: Current issues and policy implications," *Journal of Intellectual Capital,* (1:4), 1996.

[2] Chen, J., Zhu, Z. and Xie, H.Y. "Measuring intellectual capital: a new model and empirical study," *Journal of Intellectual Capital,* (5:1), 2004, pp. 195-212.

[3] Edvinsson, L. and Malone, M. S. *Intellectual Capital: Realizing your company's true value by finding its hidden brainpower.* Harper - Collins Publishers: New York, 1997.

[4] Roos, J., Roos, G., Edvinsson, L., and Dragonetti, N.C. *Intellectual capital. Navigating in the new business landscape.* New York University Press: New York, 1998.

[5] Srivihok, A. and Intrapairote, A.. Intellectual Capital Measurement: Case studies of SMEs in Thailand. Proceedings of the International E-Business. Bangkok. 2006

[6] Sullivan, P. H. *Value-driven intellectual capital: How to convert intangible corporate assets into market value*. John Wiley and Sons: New York, 2000.

[7] Sveiby, K. E. *The new organisational wealth: Managing & measuring knowledge-based assets.* Berrett-Koehler Publishers: San Francisco, 1997.

[8] Sveiby, K. E. *Methods for measuring intangible assets [Online]*.Available: http://www.sveiby.com/articles/IntangibleMethods.htm , January 2, 2004.

[9] WEKA, *http://www.cs.waikato.ac.nz/ml/weka*, 15 April 2008.

[10] Witten, I. H., and Frank,,E. *Data Mining: Practical Machine Learning Tools and Techniques*, Second Edition, Morgan Kaufmann, San Francisco, 2005.