International DSI / Asia and Pacific DSI 2007 Full Paper (July, 2007)

Classification algorithms for knowledge mapping of expert in Energy industry

Budsakorn Sukontrawongsarote[1], Anongnart Srivihok[2]

[1] Department of Computer Science, Faculty of Science,
Kasetsart University, Bangkok, Thailand
(g4764230@ku.ac.th)
[2] Department of Computer Science, Faculty of Science,
Kasetsart University, Bangkok, Thailand
(anongnart.s@ku.ac.th)

Abstract

Data classification is one technique of data mining using for creating knowledge map. The knowledge map describes who has what tacit knowledge and where the knowledge is collected. Knowledge map further helps employees to learn the jobs and expertise in organization. In many organizations there is a lack of guidelines for knowledge management and knowledge mapping. The objectives of this paper is to propose a classification algorithm for creating knowledge map of experts in an organization. The data set using in this paper are in the domain of energy experts. The performances of four classification algorithms are measured by comparing their prediction powers on expert classification. Four candidate algorithms include two types of Decision Trees (ID3 and C4.5), and two Rule-based (OneR and Prism). Results show that C4.5 algorithm is the best one, while ID3 is the last one in classification for knowledge mapping. Future research and implication are also suggested.

1. Introduction

The knowledge based economy is based on the availability of information and knowledge, to create innovative products and services. Traditional industries have been forced to employ abstract knowledge in a creative and innovative way to quickly and critically evaluate existing practices, gain insights from those practices and make new innovations. Knowledge management (KM) is a concept that articulates many aspects that characterize post-industrial production in its concern with an organization's ability to create and utilize knowledge. At present, organizations face problems and difficulties in their working process; occasionally experts are not located to solve problems in time. This is due to the lack of knowledge mapping on experts in the organization. Data mining techniques can be used to find potential useful knowledge, such as patterns and rules [15]. Also data mining tasks include clustering, classifying and association rules [16]. Classification is the techniques using for data classification by using decision trees or rule based algorithms

Definitions of the terms knowledge map or knowledge mapping is proposed by Vail[9], as follows
"A knowledge map is a visual display of captured in formation and relationships, which enables the efficient communication and learning of knowledge by observers with differing backgrounds at multiple levels of detail. The individual items of knowledge included in such a map can be text, stories, graphics, models, or number. " and also "Knowledge mapping is defined as the process of associating items of information or knowledge (preferably visually) in such a way that the mapping itself also creates additional knowledge."

Knowledge map provides understanding and using the knowledge available in an organizational setting to an individual employee, a team or an organization unit [10]. Further, it provides the relations of the people within an organization. We can find out interest areas, documents, repositories, areas of expertise, and characterize how work is being addressed within actual organization (i.e., who, what, where and when) with this knowledge map. Eppler [10] seeks to establish the conceptual and empirical basis for an innovative instrument of corporate. The knowledge maps include 5 types that can be used in managing organizational knowledge. They are knowledge-sources, - assets, -structures, -applications, and –development maps. Burkhard el at. [18] proposed a framework derived from three case studies on Knowledge Maps in

Organizations. In organizations speed, clarity, and effectiveness are essential for the transfer to knowledge. Eppler [22] proposed a simple knowledge map type based on these primary classification principles which are by purpose, by graphic form, by content, by application level and by creation mode.

Creation and maintenance knowledge map are proposed functions by utilizing information retrieval and data mining techniques Lin and Hsueh [21]. They use hierarchical clustering and k-means clustering to create knowledge map and classify new documents in to existing clusters by comparing them with the document vector and cluster centroid vector. With the Euclidean distance calculation of the two vectors. The performance of the knowledge map creation method is measured by precision and recall. Analyze of the knowledge map maintenance by applying the cross-validation method. Soman and Bobbie [20] used machine learning schemes, OneR, J48 and Naïve Bayes to classify arrhythmia for ECG medical dataset. The experimental settings used in WEKA (Waikato Environment for Knowledge Analysis) as a tool for classification. The precision in prediction of this study testing is based on 10-fold cross validation. The highest accuracy, J48 compared with the training data itself. Despite the high accuracy rate of J48, the accuracy curve is unstable when the data is spilt into training and test, whereas OneR and Naïve Bayes show stable accuracy for the same dataset. The accuracy rate of OneR is the lowest among the three algorithms. Holmes and Trigg [7] used a diagnostic tool for comparison of tree-based supervised classification model. The decision trees produced by C4.5 with default setting, although the differences produced by other options (such as making nominal attribute). The method is adapted from work on approximate tree matching and applied to decision tree. Result from the study seeks to enhance to classification accuracy of learning algorithm by using ten-fold cross-validation. The experiments show that there is a fairly strong correlation between the relative edit distance to the full tree and the cross validation error except where highly relevant nominal attributes contain many values.

Past studies have applied classification algorithms to create knowledge map in organization. However, there is scanty study in creating an expert knowledge map by using classification algorithms in Thailand. The purpose of this study is to identify a classification algorithm of knowledge mapping. It is conducted by comparing the performances of four classification algorithms. They include two Decision Trees : ID3 and C4.5 and two Rule-based : OneR and Prism algorithms.

The paper is divided to five parts. Section 2 includes the fundamental theory of classification algorithms. Section 3 is the Study Framework. The experimental results are depicted in Section 4. Finally, the last section is Conclusions and Future work.

2. Classification Algorithms

Classification is one of data mining algorithms which is based on supervised learning. It is suited for predicting or describing data set with binary or nominal categories [14]. The objective of classification is to reduce the detail and diversity of data and resulting information overwork by grouping similar data. A classification model can be used to predict the class label of unknown instants. The major classification approaches consists of decision tree, decision rules, k-nearest neighbors, Bayesian approaches, neural networks, regression-based methods and vector-based method [11]. In this section we describes theory of decision trees: ID3, C4.5 and decision rules sometimes called rule-based which are algorithms used in this study.

2.1 Decision Tree

Decision tree is a popular structure for supervised learning. It is a method for approximating discrete-value functions that is robust to noisy data and capable of learning disjunctive expression. A family of decision tree that includes widely used algorithms such as ID3, C4.5 and ASSISTANT [5].

Decision trees classify instances by sorting them down the tree from the root to some leaf node, which provides the classification of the instance. Each node in the tree specifies a test of some attribute of the instance, and each branch descending from that node corresponds to one of the possible values for this attribute. An instance is classified by starting at the root node of the tree, testing the attribute specified by this node, then moving down the tree branch corresponding to the value of the attribute. This process is then repeated for the sub tree rooted at the new node. Above these is the principle in building tree [5]. In this study we chose two algorithms to build decision tree. ID3 and C4.5 are two popular decision tree algorithms which used in this study.

(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 razo : it prefers smaller decision trees, and is therefore a heuristic [5]. Occam's razor is formalized using the concept of information theory, called entropy, which characterizes the impurity of an arbitrary collection of examples.

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. [This defines what is a Decision Tree.]
- In the decision tree at each node should be associated the non-categorical attribute which is most informative among the attributes not yet considered in the path from the root. [This establishes what is a "Good" decision tree.]
- Entropy is used to measure how informative is a node. [This defines what we mean by "Good". By the way, this notion was introduced by Claude Shannon in Information Theory.]

```
function ID3 (R: a set of non-categorical attributes,
              C: the categorical attribute,
              T: a training set) returns a decision tree;
  begin
      If T is empty, return a single node with value Failure;
      If T consists of records all with the same value for
         the categorical attribute,
         return a single node with that value;
      If R is empty, then return a single node with as value
         the most frequent of the values of the categorical attribute
         that are found in records of T; [note that then there
         will be errors, that is, records that will be improperly
         classified];
      Let D be the attribute with largest Gain(D,T)
         among attributes in R;
      Let {dj| j=1,2, .., m} be the values of attribute D;
      Let {Tj| j=1,2, .., m} be the subsets of T consisting
         respectively of records with value dj for attribute D;
      Return a tree with root labeled D and arcs labeled
         d1, d2, .., dm going respectively to the trees
         ID3(R-{D}, C, T1), ID3(R-{D}, C, T2), .., ID3(R-{D}, C, Tm);
  end ID3;
```

(2) C4.5 Algorithm

C4.5 algorithm is Quinlan's extension of his own ID3 algorithm for generating decision tree Larose [16]. This algorithm recursively visit 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 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.

Tree-Construction Algorithm

The C4.5 algorithm constructs the decision tree with a divide and conquer strategy. In C4.5, each node in a tree is associated with a set of cases. Also, case are assigned weights to take into account unknown attribute values. Initiation, the root is present and associated with whole training set and with all case weights equal to 1.0. At each node, the following divide and conquer algorithm see Program 1 is executed, trying to exploit the locally best choice, with no backtracking allowed [12].

Program 1 : Pseudocode of the C4.5 Algorithm

```
FormTree(T)
(1)   ComputeClassFrequency(T);
(2)   if OneClass or FewCases
          return a leaf;
      Create a decision node N;
(3)   ForEach Attribute A
          ComputeGain (A);
(4)   N.test = AttributeWithBestGain;
(5)   if N.test is continuous
         find Threshold;
(6)   ForEach T' in the splitting of T
(7)   if T' is Empty
         Child of N is a leaf
      else
(8)     Child of N = FormTree (T');
(9)   ComputeErrors of N;
return N
```

Figure 1. reveals example of decision tree of expert classification.

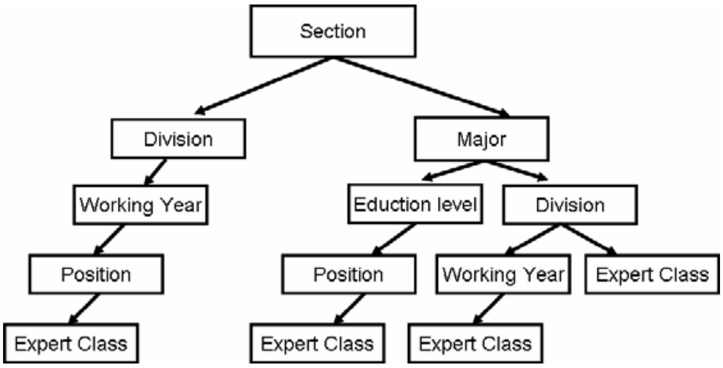

Fig. 1 Example of Decision Tree for Expert Classification

Thomassey and Fiordaliso [19] suggest that the splitting node strategy is based on the computation of the information gain ratio. Each node should hold a question relating to the attribute which is the most informative involving the set of attributes not yet measured in the path from the root to that node. Information value also called Entropy measures. The information gain associated with an attribute is computed as the difference between the information values of a node with or without the attribute. For decision tree induction rationale, the classical overfitting problem can be solve problem by pruning strategies.

Both ID3 and C4.5 apply entropy measures their splitting functions, however C4.5 has more advantages than the former since it has tree pruning function and further it can be modified to handle data sets with missing values (Quinlan, 1993)

2.2 Rule-based

The rule-based system itself uses a simple technique: It starts with a rule-base, which contains all of the appropriate knowledge encoded into IF-THEN form. It called production rules or just rules. In the IF part to some action in the THEN part. A rule provides some description of how to solve a problem. Rules are relatively easy to create and understand [17].

Rules as a knowledge representation technique, any rule consists of two parts : the IF part, called the antecedent (premise or condition) and the THEN part called the consequent (conclusion or action).

The basic syntax of a rule is :

IF <antecedent>
THEN <consequent>

In general, a rule can have multiple antecedents joined by the keyword AND (conjunction), OR (disjunction) or a combination of both.

The OneR and Prism Algorithms are interesting in make rules and easy to understand it.

(1) OneR Algorithm

The one-attribute-rule algorithm that generates a one-level decision tree proposed by Holt [25]. Each attribute value will be determined. OneR algorithm creates one rule for each attribute in the training data. The rule with the smallest error rate selected.

The algorithms is as follows :

For each value V of that attribute, create a rule :

For each attribute A:

1. count how often each class appears
2. find the most frequent class, c
3. make a rule "if A=V then C=c"

Calculate the error rate of this rule

(2) Prism Algorithm

Hong and Tseng [6] apply Prism algorithm which has the idea of information gain instead of entropy as ID3. Attribute valued pairs in terms of information theory, can be thought of discrete messages. The amount of information gain about an event in a message I is defined as :

$$I(i) = \log_2 \quad = \frac{\text{probability of event after the message is received}}{\text{probability of event before the message is received}}$$

Information gain is chosen for describing a class with a larger priority. The task of the Prism algorithm is to find the selector α x which contributes the most information gain about a specified classification δn

If the training set contains instances of more than one class, then for each class δn, Prism performs the following steps in turn .

1. Calculate the probability of occurrence, p(δn| α x), of the classification δn for each selector α x
2. Select the α x for which p(δn| α x) is a maximum then create a subset of the training set
3. Repeat steps 1 and 2 for this subset until it contains only instances of class δn.
4. The complex rule is conjunction of all the selectors used in creating the similar subset.
5. At training set, erase all instance covered by complex rule
6. Repeat steps 1-5 until all instances of class δn have been removed.

2.3 Cross-Validations

Cross validation is a method for estimating the true error of a model. When a model is built from training data, the error on the training data is a rather optimistic estimate of the error rates the model will achieve on unseen data. The aim of building a model is usually to apply the model to new, unseen data [24]. An alternative to random sub sampling is cross-validation. In this approach, each record is used the same number of times for training and exactly one for testing. This method, we partition the data into ten equal-size subsets. First, we choose nine of the subsets for training and other for testing. This approach is called a ten-fold cross-validation. The k-fold cross-validation method generalizes this approach by partition the data into k equal-sized partitions. During each run, one of the partitions is chosen for testing, while the rest of them are used for training. This procedure is repeated k times so that each partition is used for testing. The total error is found by summing up the errors for all k runs [23].

2.4 Evaluation of classification algorithms

The prediction performances of four algorithms are evaluated by using precision, recall, F-measure and Root mean-squared error. Precision and recall appropriateness have been used extensively to evaluate the retrieval performance of retrieval algorithms. However, a more careful reflection reveals problems with these two measures [4,1,2]. 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. Second, recall and precision are related measures which capture different aspects of the set of retrieved documents [8].

(1) Precision

Precision is the measurement of how much of the data returned is correct [13].

$$\text{Precision (p)} = \frac{\text{Number of correct answers predicted by system}}{\text{Number of answers given by system}}$$

(2) Recall

Recall is the measurement of how much relevant data in the system has [13].

$$\text{Recall (r)} = \frac{\text{Number of correct answers predicted by system}}{\text{Total number of possible correct answers}}$$

(3) F-measure

Precision and Recall stand in opposition to one another [13]. As precision goes up, recall usually goes down. The F-measure combines the two values.

$$\text{F-measure} = \frac{(B^2+1)*(p * r)}{B^2*(p + r)}$$

Given $B = 1$, when precision and recall are weighted equally.
B is > 1, when precision is favored.
B is < 1, when recall is favored.

(4) Root mean-squared error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} (y_t - \hat{y}_t)^2}$$

y_t is the actual value

$\hat{y}_t$ is the forecast value

n is total number of sample

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 dimensionality as the actual and predicted values. The small values of RMSE means the better power of prediction [26].

3. Study Framework

The study framework in this paper consists of four stages as following:

Stage 1 : Data Preprocessing

The preprocessing process consists of two methods : data cleaning attributes and data transformation. Data cleaning, which consists of identifying the data to be mined, then choosing appropriate input attributes and output information to represent the task. Data transformation, include organizing data in desired ways converting one type of data to another (e.g., from symbolic to nominal, numerical) defining new attributes, reducing the dimensionality of the data, removing noise, “outliers”.

Stage 2 : Classification

Classification is the prediction of nominal (discrete) values. Rules are generated from trained data and then applied to new data. It was decided to concentrate on an algorithm for generating four classification algorithms (i.e., Trees such as ID3, J4.5., Rules such as OneR, Prism). All are well supported by the text and other supplements.

Stage 3 : K-Fold Cross Validation

One crucial stage where comparison of models is using form of cross validation. This stage is consisted of training set and test set data. This paper conducts the comparison of trees-based supervised classification algorithm and rules-based. Cross validation as the method of choice for evaluation. The method of deriving specific attributes and procedures that seek to enhance the classification accuracy of a learning algorithm. The emerging standard in machine learning for estimating the error rate is to use stratified 10-fold cross validation. The data is divided randomly into ten parts, in each of which the class is represented in roughly the same proportion as in the entire dataset. Each of the ten parts is held out in turn while the learning scheme builds a model from the remaining nine parts [7].

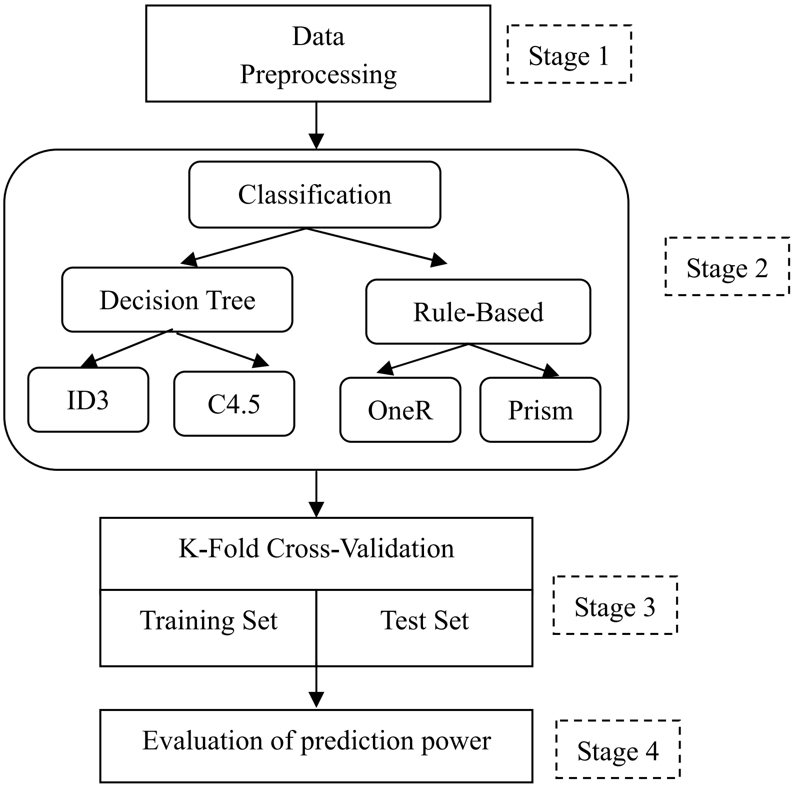

Fig. 2 Study Frame work

Stage 4 : Evaluation

There are four measurements used in this study: Precision, Recall, F-measure, Root mean-squared error. Precision show that accuracy of prediction, it should be high values. Recall opposite precision values. F-measure, if higher it show that the results have accuracies. Finally, root mean-square error, it ought to be a small number.

4. Experiment

In this section, we compare four classification algorithms such as ID3, C4.5, OneR and Prism by using 10-fold Cross-Validation. In the results of experiment of each algorithm are shown in Table 4 – Table 5. The performance measurements are Precision, Recall, F-measure and Root mean square error of k-fold corrected results.

(1) The dataset

All data used in this experiment are collected from the energy agency responsible for generating and transmitting electrical power to meet the demand as well as maintaining and inspecting electrical equipment in good condition ready for use. This responsibility can be classified by type of machine and equipment, for example, operational work, turbine, instrument and control, water system, boiler, electrical system, lignite and ash conveyor, and administration and planning. Key personnel are engineer, technician, occupational worker and administrative officer. Therefore, the classification of the experts is composed of 10 areas. Also the dataset are grouped into ten classes. The input dataset used in the Waikato Environment for Knowledge Analysis (WEKA) program, it has format extension '.arff' file [27]. The dataset has nine attributes consists of nominal attributes, since a nominal attribute can have many values. There are 343 instances, and as indicated above, in 10 classes.

Table 1. Expert Profile

Working year	Position	Education level	Faculty graduated	Major	Department	Division	Section	Expert Class
26	Tech	Voc	FBlank	Ele	DepProEleMM	Pro1	Pro1Walka1	Turbine
31	Tech	Hvoc	Fblank	Ele	DepMN	GMNele2	GMNele2ConOX2	Control_Instrument
27	Tech	Hvoc	Fblank	Ele	DepProEleMM	FuelW	Wequip3	Water_Treatment
21	Vocat	Voc	Fblank	Computer	AssEle2	DivBlank	AssEleIT	ICT
5	Arch	Bac	Farch	BrBlank	DepProEleMM	Gcivil	SecBlank	Civil
22	Spworker	Lvoc	Fblank	BrBlank	DepMN	GMNele2	GMNele2Boil2	Boiler
22	Eng	Bac	FeduInd	Peleng	DepMN	GMNele1	GMNele1ele1	Electrical
25	Eng	Bac	FeduInd	Pele	DepProEleMM	EraOxid	Woxid1	FGD
16	Eng	Bac	Feng	Meeng	DepMN	GMNcenter	SecBlank	Coal_Ash
11	Eng	Bac	Feng	Pele	DepMN	GMNele3	GMNele3MNE3	Management

Table 1 shows the details of expert profile with ten attributes. The attribute numbers and names of dataset are shown in Table 2.

Table 2. Attributes of Dataset

Attribute numbers	Attribute Names
#0	Working year
#1	Position
#2	Education level
#3	Faculty graduated
#4	Major
#5	Department
#6	Division
#7	Section
#8	Expert Class

Table 2 shows the attributes used in the experimental comprise of person attributes, which consists of : working year and Position. In education areas, consists of : Degree Education, Faculty Study and Branch Graduate. An institute attribute where it is under, consists of : Department, Division and Section. Finally, attribute type of experiences in working is Expert class.

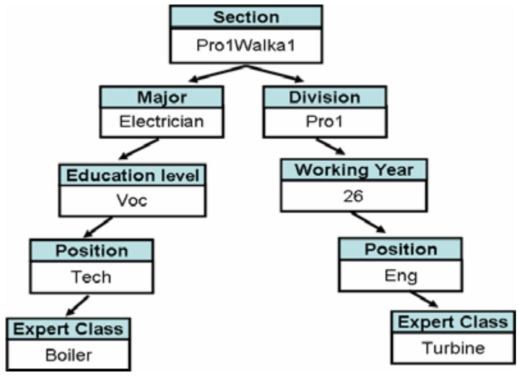

Fig. 3 Decision Tree produced by C4.5 algorithm

Figure 3 shows an example of classification by using decision tree : C4.5 algorithm. From decision tree, Section is the root node which consists of many decision nodes and leaves. A leaf node specifies a class value (such as Turbine, Control_Instrument, Boiler)

Table 3. The Rule based by Prism Algorithm

If Section = Pro1Walka1 and Major = Ele and Working year = 26 and Position = Tech and Education level = Voc and Faculty graduated = FBlank and Department = DepProEleMM and Division = Pro1 then Turbine	If Section = GMNele3MNE3 and Working year = 11 and Position = Eng and Education level = Bac and Faculty graduated = FEng and Major = PEle and Department = DepMN and Division = GMNele3 then Management

Table 3. shows an example of the classification results by using rule based: Prism algorithm.

Table 4. Comparing the prediction performances of four classifiers: ID3, C4.5, OneR and Prism

Fold	Algorithms							
	ID3		C4.5		OneR		Prism	
	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
1	64.371	19.714	77.914	22.086	76.971	**23.029**	67.057	17.657
2	64.914	19.657	78.971	21.029	77.971	22.029	67.943	18.457
3	65.543	19.286	78.886	21.114	77.657	22.343	68.486	17.143
4	65.000	20.265	79.882	20.118	77.971	22.029	68.059	18.676
5	64.412	19.265	78.853	21.147	77.559	22.441	68.353	17.941
6	66.059	17.676	79.088	20.912	79.000	21.000	68.971	**16.647**
7	65.471	18.794	**80.029**	19.971	79.088	20.912	69.029	17.206
8	65.235	19.235	79.088	20.912	78.059	21.941	67.941	18.676
9	65.412	19.088	80.588	19.412	79.324	20.676	68.706	17.559
10	66.088	18.088	79.294	20.706	77.912	22.088	69.529	17.118

Table 4 reveals the comparison of predictive accuracy on the whole dataset. Results of the four algorithms, C4.5 algorithm is more accurate than the other algorithms, and the percent of incorrect prediction by OneR algorithm is the highest among four algorithms.

Table 5. Calculate the average prediction performance scores of four classification algorithms: ID3, C4.5, OneR and Prism algorithms

Algorithm	Correct	Incorrect	Precision	Recall	F-measure	Root mean-squared error
ID3	65.250	19.106	0.692	**0.676**	0.652	0.188
C4.5	**79.259**	20.740	**0.856**	0.685	**0.734**	**0.170**
OneR	78.151	**21.848**	0.646	0.734	0.662	0.207
Prism	68.407	**17.708**	0.701	0.716	0.676	0.185

Table 5 shows results of comparison on Classification Performances by averaging 10-fold cross validation. From comparisons of prediction power of the dataset, C4.5 is the best performance than the other algorithms. Prism algorithm has the lowest incorrect score, while OneR has the highest. Comparing with precision, C4.5 is the highest performance, recall values usually go opposite direction with precision, ID3 has achieved the lowest recall. As for the F-measure C4.5 is the highest than the other algorithms. Lastly, Root mean-squared error, C4.5 is the lowest score means it prediction value is the closest to the actual value. From the above results, it can be concluded that C4.5 is the best classifier.

5. Conclusions and Future Work

It is accepted that knowledge is a valuable asset of organizations, Knowledge Management and mapping are critical for enterprise in the knowledgebase society. This study proposes an algorithm for knowledge mapping of experts in an organization to assist tacit knowledge visualization. The domain knowledge was classified by using two Decision Trees (ID3 and C4.5) and two Rule-based (OneR and Prism) algorithms. In practical experiments with 10-fold cross-validation, the four classification algorithms in terms of correct and incorrect predictions are calculated. The prediction performances of four classifiers are evaluated by four methods. The measurements include precision, recall, F-measure and RMSE. Results reveal that C4.5 algorithm is the best one in providing correct prediction, highest precision value, F-measure and lowest RMSE. While, Prism algorithm gives smaller percent of correct prediction, this algorithm gives the smallest number of incorrect prediction. OneR algorithm has the highest numbers for incorrect prediction, precision, F-measure and RMSE. Thus, C4.5 is proposed to be the best classifier for knowledge mapping.

Implication by relying on Java's Unicode capability, a new Java implementation of the PAT-Tree. Phrase Extraction has been completed, and new research is under way to apply this approach to Spanish, Tamil and English to determine its external validity. Furthermore, a multilingual entity extraction system

Future work is suggested by applying data mining technique such as association rules for classification of organizational knowledge mapping. Using different algorithm might result in different performance In order to increase the prediction power of classification. Clustering algorithms such as K-Mean or Self-Organization Map (SOM) might be applied to segment dataset to similar group. Then each group is used to build decision tree for knowledge classification.

6. Acknowledgement

We would like to thank the Electricity Authority of Thailand for dataset which use in this research.

References

[1] Vijay V. Raghavan, Gwang S. Jung and Peter Bollmann; A critical investigation of recall and precision as measures of retrieval system performance, ACM Transactions on Office and Information Systems, pp 205-229, 1989.

[2] Jean Tague-Sutcliffe; Measuring the informativeness of a retrieval process. In Proceeding of the 15th Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval, pp 23-36, 1992.

[3] John Ross Quinlan; C4.5: Programs for Machine Learning. Morgan-Kaufmann Publishers, San Mateo, CA, 1993.

[4] Robert Korfhage; Information Storage and Retrieval, John Wiley & Sons, Inc.,1997.

[5] Tom Mitchell; Machine Learning, The McGraw-Hill Companies, Inc, International Edition, 1997.

[6] Tzung-Pei Hong and Shian-Shyong Tseng; Two-phase PRISM Learning Algorithms, Systems, Man, and Cybernetics, Computational Cybernetics and Simulation, IEEE International Conference, Vol. 4, pp 3895 – 3899, 1997.

[7] Geoffrey Holmes and Leonard Trigg; A Diagnostic Tool for Tree Based Supervised Classification Learning Algorithms, Proceedings. ICONIP '99. 6^{th} International Conference on Vol. 2, pp 514 – 519, 1999.

[8] Ricardo Baeza-Yates and Berthier Ribeiro-Neto; Modern information retrieval, pp 81, 1999.

[9] Vail, E. F; Mapping Organizational Knowledge, Knowledge Management Review, Issue 8, May/June, pp 10-15, 1999.

[10] Martin J. Eppler; Making Knowledge Visible Through Intranet Knowledge Maps: Concepts, Elements, Cases, System Sciences, Proceedings of the 34^{th} Annual Hawaii International Conference, pp 225-444, 2001.

[11] Heide Hrucher, Gerhard Knolmayer and Marc-Andre Mittermayer, Document Classification Method for Organizing Explicit Knowledge, Third European Conference on Organizational Knowledge, Learning and Capabilities in Athens, 2002.

[12] Salvatore Ruggieri; Efficient C4.5, IEEE Transactions on Knowledge and Data Engineering, Vol. 14, March/April 2002.

[13] Miriam Butt; Information Retrieval (based on Jurafsky and Martin), 2003.

[14] Richard J. Roiger and Michael W. Geatz; Data Mining : a tutorial-based primer, 2003.

[15] Zhe Huang, Yun-Quan Hu; Applying AI Technology and Rough Set Theory to Mine Association Rules for Supporting Knowledge Management, Proceedings of the Second International Conference on Machine Learning and Cybernetics, Xi'an, IEEE, 2003.

[16] Daniel T. Larose; Discovering Knowledge in Data, Wiley-Interscience, A John Wiley & Sons, Inc., Publication 2005.

[17] Michael Negnevitsky; Artificial Intelligence A Guide to Intelligence System Second Edition, Addison Wesky, 2005.

[18] Remo Burkhard, Michael Meier, Matthias Smis, Jill Allemang, and Laura Honisch; Beyond Excel and Powerpoint : Knowledge Maps for the Transfer and Creation of Knowledge in Organizations, Proceedings of the Ninth International Conference on Information Visualisation (IV'05), 2005.

[19] Sebastien Thomassey and Antonio Fiordaliso; A hybrid sales forecasting system based on clustering and decision trees, European Journal of Operational Research, pp 518-542, 2005.

[20] Thara Soman and Patrick O.Bobbie; Classification of Arrhythmia Using Machine Learning Techniques, Proceeding of 4^{th} International Conference on System Science and Engineering (ICOSSE), 2005.

[21] Fu-ren Lin and Chih-ming Hsueh; Knowledge Map Creation and Maintenance for Virtual Communities of Practice, Information Processing and Management 42, Issue 2, pp 551-568, 2006.

[22] Martin Eppler; Toward a Pragmatic Taxonomy of Knowledge Maps : Classification Principles, Sample Typologies, and Application Examples, Proceedings of the Information Visualization (IV'06), pp195 – 204, 2006.

[23] Pang-Ning Tan, Michael Steinbach and Vipin Kumar; Introduction to data mining, Pearson Education, Inc., USA, 2006.

[24] Cross Validation : http://datamining.togaware.com/survivor/Cross_Validation.html 16 April 2007

[25] OneR : http://en.wikipedia.org/wiki/One-attribute-rule 16 April 2007

[26] RMSE : http://grb.mnsu.edu/grbts/doc/manual/Error_Measurements.html 16 April 2007

[27] WEKA : website http://www.cs.waikato.ac.nz/~ml/weka/index.html 1 March 2007