1

Data Mining Classification Techniques for Human Talent Forecasting

Hamidah Jantan[1], Abdul Razak Hamdan[2] and Zulaiha Ali Othman[2]
[1]Faculty of Computer and Mathematical Sciences UiTM, Terengganu, 23000 Dungun, Terengganu,
[2]Faculty of Information Science and Technology UKM, 43600 Bangi, Selangor, Malaysia

1. Introduction

In knowledge management process, data mining technique can be used to extract and discover the valuable and meaningful knowledge from a large amount of data. Nowadays, data mining has given a great deal of concern and attention in the information industry and in society as a whole. This technique is an approach that is currently receiving great attention in data analysis and it has been recognized as a newly emerging analysis tool (Osei-Bryson, 2010; Park, 2001; Sinha, 2008; Tso & Yau, 2007; Wan, 2009; Zanakis, 2005; Zhuang et al., 2009). Additionally, among the major tasks in data mining are classification and prediction; concept description; rule association; cluster analysis; outlier analysis; trend and evaluation analysis; statistical analysis and others. Classification and prediction tasks are among the popular tasks in data mining; and widely used in many areas especially for trend analysis and future planning. In fact, classification technique is supervised learning, which is the class level or prediction target is already known. As a result, the classification model which is represented through rules structures will be constructed in the classification process. In this case, the constructed model will be representing the precious knowledge and it can be used for future planning.

There are many areas which adapted this approach to solve their problems such as in finance, medical, marketing, stock, telecommunication, manufacturing, health care, customer relationship and etc. However, the data mining application has not attracted much attention from people in Human Resource (HR) field (Chien & Chen, 2008; Ranjan, 2008). Besides that, in our previous study, most of the prediction applications are used to predict stock, demand, rate, risk, event and others; but there are quite limited studies on human prediction. In addition prediction applications are mainly developed in business and industrious fields; and quite restricted studies involved human talent in an organization (Jantan et al., 2009). HR data can provide a rich resource for knowledge discovery and for decision support system development.

Recently, an organization has to struggle effectively in term of cost, quality, service or innovation. All these depend on having enough right people with the right skills, employed in the appropriate locations at appropriate point of time. In HR, among the challenges of HR professionals are managing an organization talent known as talent management. Talent management involves a lot of managerial decisions and these types of decisions are very uncertain and difficult. Besides that, these decisions depend on various factors such as human experience, knowledge, preference and judgment. The process to identify the existing talent in an organization is among the top talent management challenges and the important issue (A TP Track Research Report 2005). In addition, talent management is defined as an outcome to ensure the right person is in the right job (Cubbingham, 2007). Talent in an organization is evaluated based on the position that he/she holds, and the position is represented by the talent ability that he/she has. Due to those reasons, this study attempts to use classification techniques in data mining to handle issue on talent forecasting. In this study, academic talent type of data in higher learning institution has been chosen as the datasets to represent human talent. As a result, the purpose of this article is to suggest the potential classification techniques for human talent forecasting through some experiments using selected classification algorithms.

This chapter is organized as follows. The second section describes the related work on classification and prediction in data mining; researches on data mining in HR especially for talent management; and human talent forecasting using data mining technique. The third section discusses on experiment setup in this study. Next, the forth section shows experiment results and discussions. Then, section five suggests some related future works. Finally, the paper ends at Section 6 with the concluding remarks acknowledged.

2. Related work

2.1 Classification and prediction in data mining

Data mining tasks are generally categorized as clustering, association, classification and prediction (Chien & Chen, 2008; Ranjan, 2008). Over the years, data mining has evolved various techniques to perform the tasks that include database oriented techniques, statistic, machine learning, pattern recognition, neural network, rough set and etc. Database or data warehouse are rich with hidden information that can be used to provide intelligent decision making. Intelligent decision refers to the ability to make automated decision that is quite similar to human decision. Classification and prediction in machine learning are among the techniques that can produce intelligent decision. At this time, many classification and prediction techniques have been proposed by researchers in machine learning, pattern recognition and statistics.

Classification and prediction in data mining are two forms of data analysis that can be used to extract models to describe important data classes or to predict future data trends (Han & Kamber, 2006). The classification process has two phases; the first phase is learning process, the training data will be analyzed by the classification algorithm. The learned model or classifier shall be represented in the form of classification rules. Next, the second phase is classification process where the test data are used to estimate the accuracy of the classification model or classifier. If the accuracy is considered acceptable, the rules can be applied to the classification of new data (Fig. 1).

Several techniques that are used for data classification are decision tree, Bayesian methods, Bayesian network, rule-based algorithms, neural network, support vector machine, association rule mining, k-nearest-neighbor, case-based reasoning, genetic algorithms, rough sets, and fuzzy logic. In this study, we attempt to use three main classification techniques i.e. decision tree, neural network and k-nearest-neighbor. However, decision tree and neural network are found useful in developing predictive models in many fields(Tso & Yau, 2007). The advantage of decision tree technique is that it does not require any domain knowledge or parameter setting, and is appropriate for exploratory knowledge discovery. The second technique is neural-network which has high tolerance of noisy data as well as the ability to classify pattern on which they have not been trained. It can be used when we have little knowledge of the relationship between attributes and classes. Next, the K-nearest-neighbor technique is an instance-based learning using distance metric to measure the similarity of instances. All these three classification techniques have their own advantages and disadvantages, for that reasons, this study endeavor to explore these classification techniques for human talent data. Besides that, data mining technique has been applied in many fields, but its application in HR is very rare (Chien & Chen, 2008).

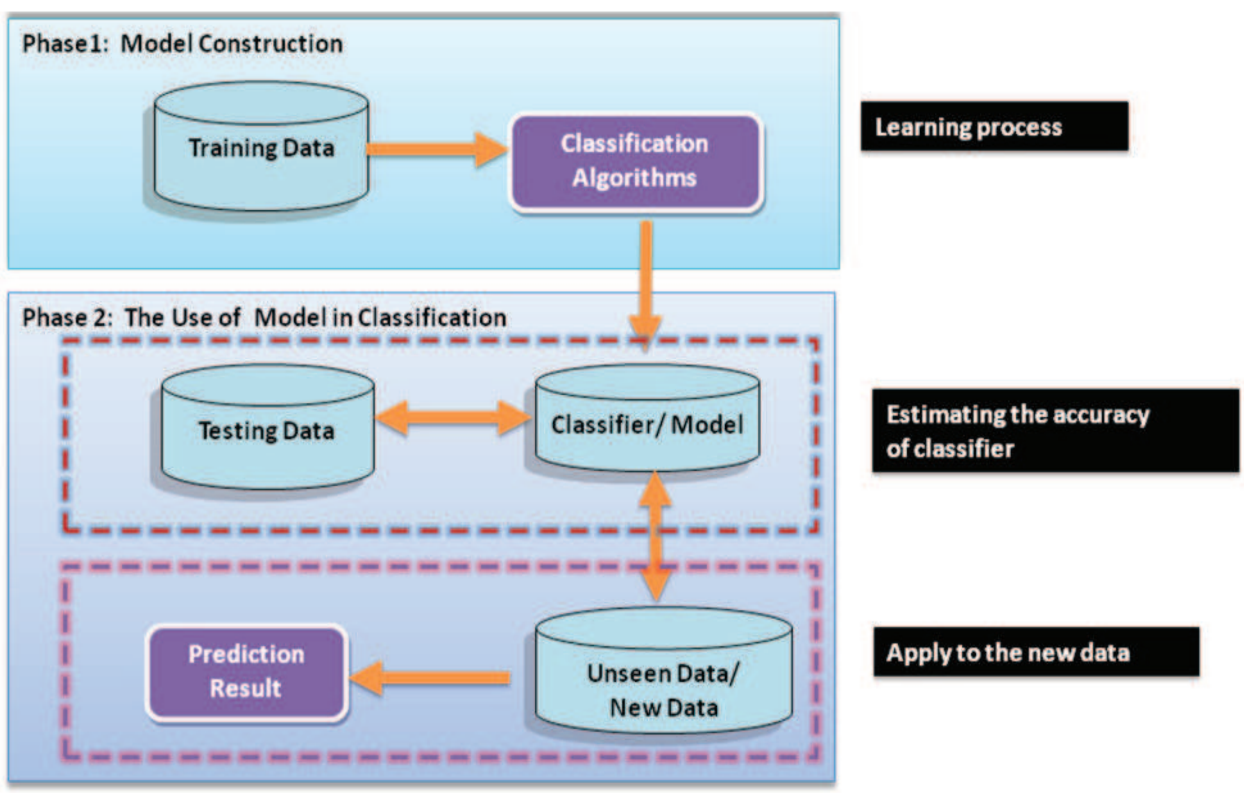

Fig. 1. Classification and Prediction in Data Mining

Recently, there are some researches that show great interest on solving HR problems using data mining approach (Ranjan, 2008). Table 1 lists some of the tasks in human resource that use data mining technique, and it shows there are quite limited studies on data mining in human resource domain. In addition, until now there are quite limited discussions on talent management such as for talent forecasting, career planning and talent recruitment use data mining approach. In HR, data mining technique used focuses on personnel selection especially to choose the right candidates for a job. The classification and prediction in data mining for HR problems are infrequent and there are some examples such as to predict the length of service, sales premium, to persistence indices of insurance agents and analyze miss-operation behaviors of operators (Chien & Chen, 2008). Due to these reasons, this study attempts to use data mining classification techniques to forecast potential employees as substantial of talent management task using the past experience knowledge.

HR Task	Data Mining Technique
Personnel selection	Decision tree (Chien & Chen, 2008), Fuzzy Logic and Data Mining (Tai & Hsu, 2005) Rough Set Theory(Chien & Chen, 2007)
Training	Association rule mining (Chen et al., 2007)
Employee Development	Fuzzy Data Mining and Fuzzy Artificial Neural Network (Huang et al., 2006) Decision Tree (Tung et al., 2005)
Performance Evaluation	Potential to use Decision Tree (Zhao, 2008)

Table 1. Data mining Techniques in HRM.

2.2 Talent management and data mining

In any organization, talent management has become an increasingly crucial approach in HR functions. Talent is considered as the capability of any individual to make a significant difference to the current and future performance of the organization (Lynne, 2005). In fact, managing talent involves human resource planning that emphasizes processes for managing people in organization. Besides that, talent management can be defined as a process to ensure leadership continuity in key positions and encourage individual advancement; and decision to manage supply, demand and flow of talent through human capital engine (Cubbingham, 2007). Talent management is very crucial and needs some attention from HR professionals. TP Track Research Report has found that among the top current and future talent management challenges are developing existing talent; forecasting talent needs; attracting and retaining the right leadership talent; engaging talent; identifying existing talent; attracting and retaining the right leadership and key contributor; deploying existing talent; lack of leadership capability at senior levels and ensuring a diverse talent pool (A TP Track Research Report 2005). The talent management process consists of recognizing the key talent areas in the organization, identifying the people in the organization who constitute key talent, and conducting development activities for the talent pool to retain and engage them and also have them ready to move into more significant roles (Cubbingham, 2007) (Fig. 2). These processes involve HR activities that need to be integrated into an effective system (CHINA UPDATE, 2007) (Fig. 2).

In this study, we focus on one of the talent management challenges i.e. to identify the existing talent regarding the key talent in an organization by predicting their performance using previous employee performance records in databases. In this case, we use the past related employee data regarding on their talent by using classification technique in data mining.

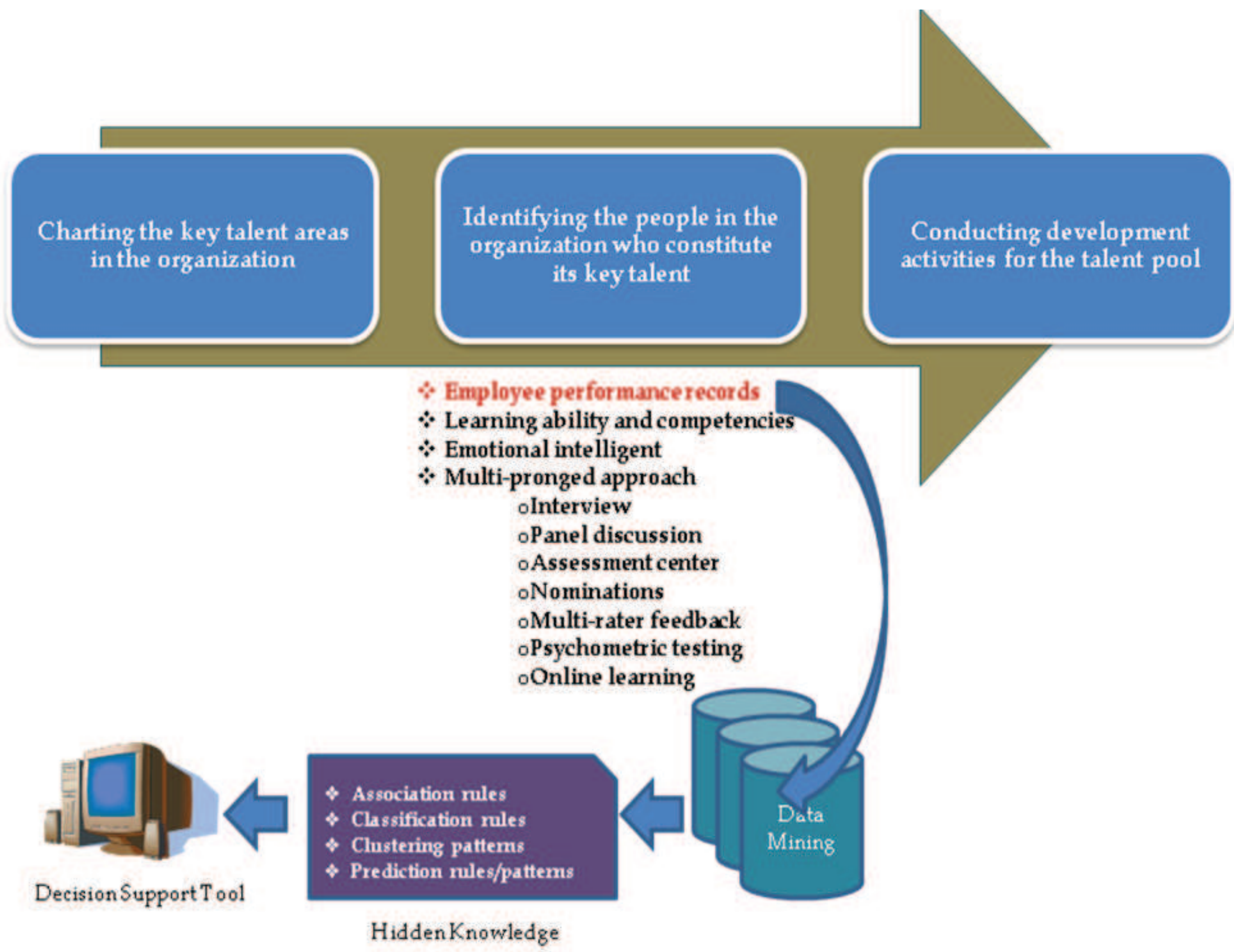

Fig. 2. Data mining and Talent Management

2.3 Human talent forecasting

Recently, with the new demand and increased visibility, HR seeks a more strategic role by turning to data mining methods (Ranjan, 2008). This can be done by discovering generated patterns as useful knowledge from the existing data in HR databases. Thus, this study concentrates on identifying the patterns that relate to the human talent. The patterns can be generated by using some of the major data mining techniques such as clustering to list the employees with similar characteristics, to group the performances and etc. From the association technique, patterns that are discovered can be used to associate the employee's profile for the most appropriate program/job, associated with employee's attitude toperformance and etc. In prediction and classification task, the pattern discovered can be used to predict the percentage accuracy in employee's performance, behavior, and attitudes, predict the performance progress throughout the performance period, and also identify the best profile for different employee and etc. (Fig. 3). The match of data mining problems and talent management needs are very crucial. Therefore, it is very important to determine the suitable data mining techniques for talent management problems.

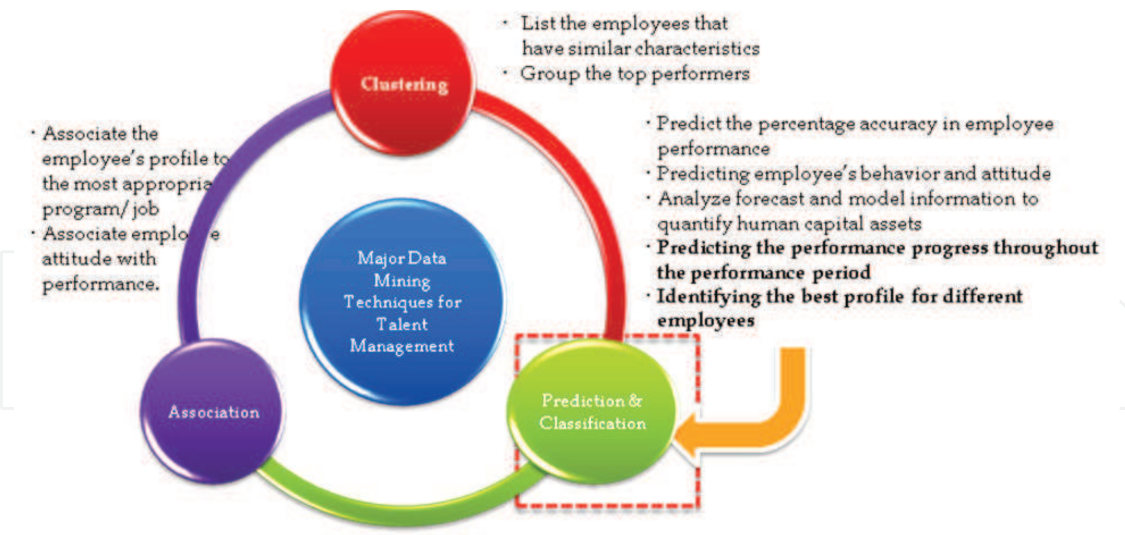

Fig. 3. Data mining Tasks for Talent Management

3. Experiment setup

This experiment attempts to propose the potential data mining classifier for human talent data. The proposed classifier can be used to generate talent performance classification patterns from employee's performance databases. Subsequently, the generated classification patterns can be employed in decision support tool for human talent prediction. The basic process for classification and prediction in data mining has been discussed in the related work (Fig. 1). The experiment setup in this study has several tasks such as simulated data construction, outlier placing, attribute reduction and accuracy of model determination as shown in Fig. 4. However, due to the difficulties to get real data from HR department, because of the confidentiality and security issues, for the exploratory purposes, this study simulates two human talent datasets

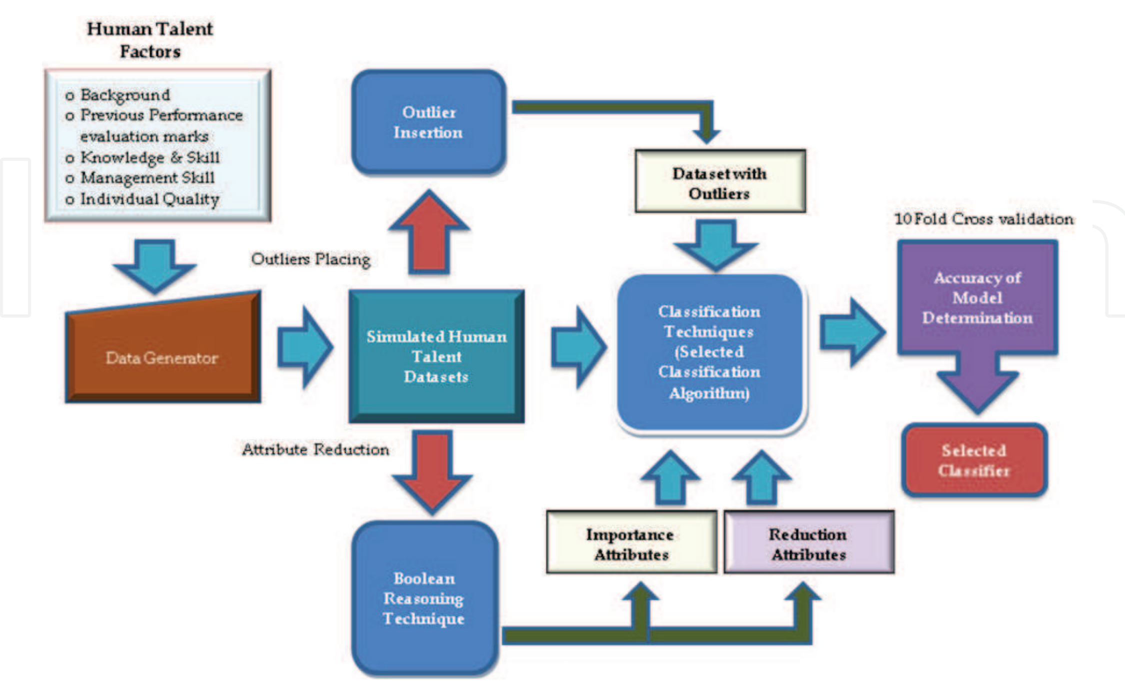

Fig. 4. Experiment Setup

using dataset rule generator shown in Table 2. The first dataset contains one hundred data (*dataset1*) and the second dataset has a thousand performance data *(dataset2)* based on human talent performance factors. In many cases, simulated or syntactic data is an ideal data and can produce a good data mining model. For that reason, in this study uses outlier placing task for *dataset1* to handle that issue and that new dataset known as *dataset3*.

In this experiment, the selected classification techniques used are based on the common techniques used for classification and prediction in data mining. As mentioned earlier in related work, the classification techniques chosen are neural network which is quite popular in data mining community and used as pattern classification technique (Witten & Frank, 2005). The decision tree known as 'divide-and–conquer' approach is from a set of independent instances for classification and the nearest neighbor is for classification that are based on the distance metric. Table 3 summarizes the selected classification techniques in data mining, such as decision tree, neural network and nearest neighbor. In this study, we attempt to use C4.5 and Random Forest for decision tree category; Multilayer Perceptron (MLP) and Radial Basic Function Network (RBFC) for neural network category; and K-Star for the nearest neighbor category.

Factor and Attributes	Rules
Background/ Demographic (D1-D8) (a1-a8) Class level – D4/a4	D1 = RANDBETWEEN (1950-1983), D2 = RANDBETWEEN (1,2,3,4), D3 = RANDBETWEEN (0,1), D4 = RANDBETWEEN ((1-4), D5= RANDBETWEEN (1975-2008) and G2 = IF (D5-D1<25 THEN D1+25 ELSE D5) I2 = G2+RANDBETWEEN(5,10) D6 = IF(I2>2008 THEN 0 ELSE I2) K2= G2+RANDBETWEEN(6,15) D7 = IF(K2>2008THEN 0 ELSE K2) M2= G2+RANDBETWEEN(10,30) D8 = IF(M2>2008 THEN 0 ELSE M2)
Previous performance evaluation (DP1-DP15) (a9-a22)	{DP1,DP15}= RANDBETWEEN (75-100)
Knowledge and skill (PQA-PQH) (a23-a42)	{ PQA,PQC1,PQC2,PQC3,PQD1, PQD2,PQD3,PQE1,PQE2,PQE, PQE4,PQE5,PQF1,PQF2,PQG1, PQG2,PQH1,PQH2,PQH3,PQH4} =RANDBETWEEN (1-10)
Management skill (PQB, AC1-AC5) (a43 –a48)	{PQB }=RANDBETWEEN(1-10) {AC1, AC2,AC3,AC4,AC5}= RANDBETWEEN (0-5)
Individual Quality (T1-T2, SO, AA1-AA2) (a49-a53)	{T1,T2} = RANDBETWEEN (1-10) {SO,AA1,AA2} = RANDBETWEEN (0-5)

Table 2. Rules to Generate Simulated Dataset

Data Mining Techniques	Classification Algorithm
Decision Tree	• *C4.5* (Decision tree induction – the target is nominal and the inputs may be nominal or interval. Sometimes the size of the induced trees is significantly reduced when a different pruning strategy is adopted). • *Random forest* (Choose a test based on a given number of random features at each node, performing no pruning. Random forest constructs random forest by bagging ensembles of random trees).
Neural Network	• *Multi Layer Perceptron* (An accurate predictor for underlying classification problem. Given a fixed network structure, we must determine appropriate weights for the connections in the network). • *Radial Basic Function Network* (Another popular type of feed forward network, which has two layers, not counting the input layer, and differs from a multilayer perceptron in the way that the hidden units perform computations).
Nearest Neighbor	• *K*Star* (An instance-based learning using distance metric to measure the similarity of instances and generalized distance function based on transformation

Table 3. Selected Classification Algorithm

The human talent factor in this case study is for academic talent in higher learning institution. The academic talent factors are extracted from the common practice for evaluation, performance evaluation documents and expertise experiences. Besides the human performance factors, the talent background and management skill are also considered in the process to identify the potential talent. In this experiment, the training dataset contains 53 related attributes from five performance factors demonstrated in Table 4. The target class for the dataset is the academic position (*D4*) which is representing as professor, associate professor, senior lecturer and lecturer. The classification technique used is based on 10 fold cross validation training and test dataset. In this experiment, the data mining tools used are WEKA and ROSETTA toolkit. This experiment has two phases; the first phase is to identify the possible techniques using selected classifier algorithm for full attributes of data. In this case, we use all the attributes which are defined before for the full dataset.

Besides that, this experiment concentrates on the accuracy of selected classifiers in order to identify potential classifier algorithm for the datasets. The accuracy of classifier is based on the percentage of test set samples that are correctly classified. The second phase of experiment is to compare the accuracy of classifier for attribute reduction. In this case, Boolean reasoning technique is used to select the most relevant or important attributes from the dataset. The attribute reduction phase is divided into two stages. The first stage is attribute reduction using the shortest length attribute, which is used by many researches in attribute reduction process. The aim of this process is to determine the important attributes for the data set, which is known as attribute reduction dataset (AR). The second stage is for

Factor and Attributes	Variable Name	Meaning
Background (7)	D1,D2,D3,D5,D6, D7,D8	Age ,Race, Gender, Year of service, Year of Promotion 1, Year of Promotion 2, Year of Promotion 3
Previous performance evaluation (15)	DP1,DP2,DP3, DP4,DP5,DP6, DP7,DP8,PP9, DP10, DP11,DP12, DP13,DP14, DP15	Performance evaluation marks for 15 years
Knowledge and skill (20)	PQA,PQC1,PQC2, PQC3,PQD1, PQD2,PQD3,PQE1, PQE2,PQE, PQE4,PQE5,PQF1, PQF2,PQG1, PQG2,PQH1,PQH2, PQH3,PQH4	Professional qualification (Teaching, supervising, research, publication and conferences)
Management skill (6)	PQB,AC1,AC2,AC3,AC4,AC5	Student obligation and administrative tasks
Individual Quality (5)	T1,T2,SO,AA1,AA2	Training, award and appreciation

Table 4. Factors and Attributes for Academic Talent

the combination of important attribute which is known as importance attributes dataset (IA). In this case, we attempt to study the accuracy of the classifier using all importance attributes. Finally, the experiment results for each phase is evaluated using the statistical significant test in order to determine the most significant classifier for each of datasets and it will be considered as the potential classifier for human talent data.

4. Result and discussion

In this experiment, the accuracy of classification techniques is based on the selected classifier algorithm. In the first phase, the accuracy for each of the classifier algorithm for full attributes for three datasets is shown in Table 5. The results for full attribute present the highest accuracy of model is C4.5 (95.14%, 99.90% and 90.54%) which is the results could be considered as an indicator to the potential classification algorithm for human talent data (Fig. 5.).

Classification Algorithm	Dataset1	Dataset2	Dataset3
C4.5	95.14	99.90	90.54
Random forest	74.91	95.43	71.80
Multi Layer Perceptron (MLP)	87.16	99.84	84.55
Radial Basis Function Network	91.45	99.98	87.09
K-Star	92.06	97.83	87.79

Table 5. Accuracy of Model for Full Attributes

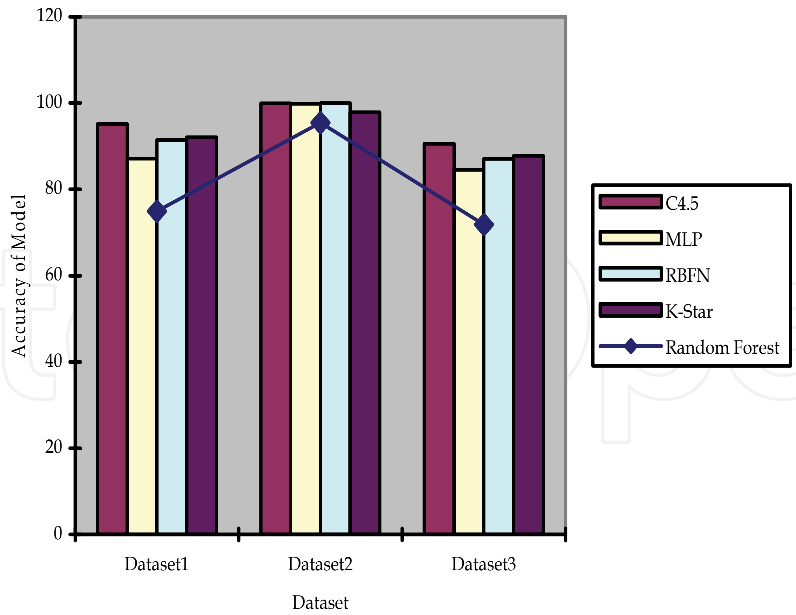

Fig. 5. Accuracy of Model for Full Attributes

The result for full attributes shows us the more data that we used (*dataset2*) in training process the highest accuracy of model can be developed. Besides that, the accuracy for dataset3 which contains outliers is slightly down for all classifiers, this result demonstrates the effect of outliers in dataset for accuracy of the model. The second phase of the experiment is considered as a relevant analysis process in order to determine the accuracy of the selected classification technique using datasets with attribute reduction. In this experiment, we focus on dataset1 and dataset2. The purpose of attribute reduction process is to select the most relevant attribute in the dataset. The reduction process is implemented using Boolean reasoning technique. Through attribute reduction, we can decrease the preprocessing and processing time and space. Table 6. shows the relevant analysis results for attribute reduction, five (5) attributes are selected, all the attributes are from the background factor. By using these attributes reduction variables, the second phase of experiment is implemented. The aim of this experiment is to find out the accuracy of the classification techniques with attribute reduction using the shortest length attributes and combination of the important attributes after reduction process.

Variable Name	Meaning
D1,D5,D6,D7,D8	Age, Year of service, Year of Promotion 1, Year of Promotion 2, Year of Promotion 3

Table 6. Important Attributes from Atribut Reduction

Table 7. shows the accuracy of the classification algorithm with attribute reduction for the shortest length methods (AR dataset). The C4.5 classifier has the highest percentage of accuracy in the first stage of second phase experiment (Table 7.) but the accuracy has declined at this stage.

Classification Algorithm	Dataset1	Dataset2
C4.5	61.06	63.21
Random forest	58.85	62.49
Multi Layer Perceptron (MLP)	55.32	60.16
Radial Basis Function Network(RBFN)	59.52	64.05
K-Star	60.22	63.92

Table 7. Accuracy of Model for Attribute Reduction

In this experiment, the result indicates more attributes used in dataset that will affect the accuracy of the classifier. Consequently, this result illustrates most of the attributes in dataset are important and should be considered. However, with the combination of attributes from reduction process (IA dataset) in the second stage of experiment, the accuracy of classifier is higher compared to the shortest length attributes (AR dataset). Table 8. shows the accuracy of classifier for importance attributes for *dataset1* and *dataset2*. The C4.5 classifier has the highest accuracy for both datasets at this stage of experiment. Fig. 6. shows the accuracy of model for AR datasets and IA datasets in the second phase experiment.

Classification Algorithm	Dataset1	Dataset2
C4.5	95.63	99.89
Random forest	86.50	99.88
Multi Layer Perceptron (MLP)	79.49	99.91
Radial Basis Function Network(RBFN)	84.41	99.96
K-Star	78.40	99.95

Table 8. Accuracy of Model for Importance Attribute

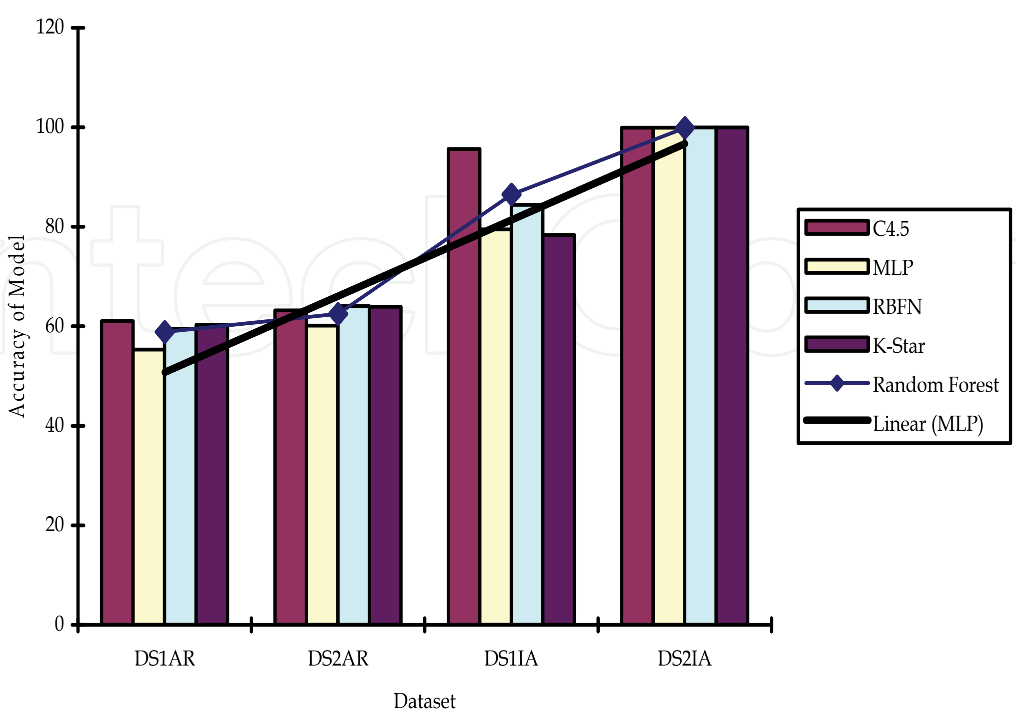

Fig. 6. The Accuracy of Model for Attribute Reduction and Importance Attributes

Consecutively, to propose the potential classifier for human talent data, the statistical significant test is conducted using t-test evaluation. By using the pair t-test as shown in Table 9, a positive mean difference in accuracy shows that the C4.5 has the highest value of positive mean which is significantly better than other classifiers. For the accuracy criterion, C4.5 is significantly better than Random Forest and MLP, with a p-value < 0.05. In addition, decision tree can produce a model which may represent interpretable rules or logic statement and can be performed without complicated computations and the technique can be used for both continuous and categorical variables. This technique is more suitable for predicting categorical outcomes and less appropriate for application to time series data (Tso & Yau, 2007). Besides that, the decision tree classifiers are a quite popular technique because the construction of tree does not require any domain knowledge or parameter setting, and therefore is appropriate for exploratory knowledge discovery.

	Paired Samples	Mean	SD	t	df	p-value
Pair 1	C4.5 – Random Forest	7.93000	8.45564	2.481	6	*0.048
Pair 2	C4.5 - MLP	5.56286	5.56322	2.646	6	*0.038
Pair 3	C4.5 - RBFN	2.70143	4.15154	1.722	6	0.136
Pair 4	C4.5 - KStar	3.60000	6.17387	1.543	6	0.174

SD: Standard Deviation; t: significant ratio; df: degrees of freedom; p: significant 2-tailed value;
* most significant

Table 9. Pair T-Test Result on Accuracy of Model for C4.5

In these experiments, we observe the great potential to use C4.5 classification algorithm in the next stage of data mining process i.e. prediction using the constructed classification model. Besides that, these results also show about the suitability of C4.5 classifier for the human talent datasets.

5. Future works

In this study, due to the difficulties to obtain human talent data, we have to simulate the data for exploratory purposes and setup the classification experiment using the data. In this case, knowledge discovered or constructed classification model by using the proposed classifier for the datasets cannot be used to represent the real problems. In future works, the similar experiment setup can be applied to the real data in order to use classification model constructed by the proposed classifier. Besides that, other Data mining techniques such as Support Vector Machine (SVM), Fuzzy logic and Artificial Immune System (AIS) should also be considered for future work on classification techniques using the same dataset.
In some cases, the attribute relevancy has also become a factor on the accuracy of the classification algorithm. In the next experiment, the attribute reduction process should be applied to other reduction techniques in order to confirm these findings whether the number of attributes will affect the accuracy of the classifier. Besides that, the C4.5 classifier has the highest accuracy in the experiment; the accuracy for other decision tree classifier also needs to be experimented in order to validate these findings.

6. Conclusion

This article has described the significance of the study using data mining for talent management especially for classification and prediction. However, there should be more data mining techniques applied to the different problem domains in HR field of research in order to broaden our horizon of academic and practice work on data mining in HR. In addition, C4.5 classifier algorithm is the potential classifier in this experiment. Thus, this technique can be used for real human talent data in the next prediction phase i.e classification rules construction. These generated classification rules can be used to predict the potential talent for the specific task in an organization. In HRM, there are several tasks that can be solved using this approach, for examples, selecting new employees, matching people to jobs, planning career paths, planning training needs for new and senior employee, predicting employee performance, predicting future employee and etc. In conclusion, the ability to continuously change and obtain new understanding about classification and prediction in HR field has thus, become the major contribution to HR data mining.

7. References

A TP Track Research Report (2005). *Talent Management: A State of the Art*: Tower Perrin HR Services.

Chen, K. K., Chen, M. Y., Wu, H. J., & Lee, Y. L. (2007). *Constructing a Web-based Employee Training Expert System with Data Mining Approach*. Paper presented at the Paper in The 9th IEEE International Conference on E-Commerce Technology and The 4th IEEE International Conference on Enterprise Computing, E-Commerce and E-Services (CEC-EEE 2007).

Chien, C. F., & Chen, L. F. (2007). Using Rough Set Theory to Recruit and Retain High-Potential Talents for Semiconductor Manufacturing. *IEEE Transactions on Semiconductor Manufacturing, 20*(4), 528-541.

Chien, C. F., & Chen, L. F. (2008). Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry. *Expert Systems and Applications, 34*(1), 380-290.

CHINA UPDATE. (2007). HR News for Your Organization : The Tower Perrin Asia Talent Management Study. Retrieved from www.towersperrin.com. 7/1/2008.

Cubbingham, I. (2007). Talent Management : Making it real. *Development and Learning in Organizations, 21*(2), 4-6.

Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kaufmann Publisher.

Huang, M. J., Tsou, Y. L., & Lee, S. C. (2006). Integrating fuzzy data mining and fuzzy artificial neural networks for discovering implicit knowledge. *Knowledge-Based Systems, 19*(6), 396-403.

Jantan, H., Hamdan, A. R., & Othman, Z. A. (2008). *Data Mining Techniques for Performance Prediction in Human Resource Application.* Paper presented at the 1st Seminar on Data Mining and Optimization, Selangor.

Jantan, H., Hamdan, A. R., & Othman, Z. A. (2009, 25-27 February 2009). *Knowledge Discovery Techniques for Talent Forecasting in Human Resource Application.* Paper presented at the World Academy of Science, Engineering and Technology, Penang, Malaysia.

Lynne, M. (2005). *Talent Management Value Imperatives : Strategies for Execution*: The Conference Board.

Osei-Bryson, K.-M. (2010). Towards supporting expert evaluation of clustering results using a data mining process model. *Information Sciences, 180*(3), 414-431.

Park, S. C., Piramuthu, S., & Shaw, M.J. (2001). Dynamic rule refinement in knoledge-based data mining systems. *Decision Support System, 31*(2), 205-222.

Ranjan, J. (2008). Data Mining Techniques for better decisions in Human Resource Management Systems. *International Journal of Business Information Systems, 3*(5), 464-481.

Sinha, A. P., & Zhao, H. (2008). Incorporating domain knowledge into data mining classifiers: An application in indirect lending. *Decision Support System, 46*(1), 287-299.

Tai, W. S., & Hsu, C. C. (2005). A Realistic Personnel Selection Tool Based on Fuzzy Data Mining Method. Retrieved from www.atlantis-press.com/php/download_papaer?id=46, 9/1/2008.

Tso, G. K. F., & Yau, K. K. W. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy, 32*, 1761-1768.

Tung, K. Y., Huang, I. C., Chen, S. L., & Shih, C. T. (2005). Mining the Generation Xer's job attitudes by artificial neural network and decision tree - empirical evidence in Taiwan. *Expert Systems and Applications, 29*(4), 783-794.

Wan, S., & Lei, T.C. (2009). A knowledge-based decision support system to analyze the debris-flow problems at Chen-Yu-Lan River, Taiwan. *Knowledge-Based System, 22*(8), 580-588.

Witten, I. H., & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco: Morgan Kaufmann Publisher.

Zanakis, S. H., Fernandez, I.B. (2005). Competitiveness of nations: A knowledge discovery examination. *European Journal of Operational Research, 166*(1), 185-211.

Zhao, X. (2008). *An Empirical Study of Data Mining in Performance Evaluation of HRM.* Paper presented at the International Symposium on Intelligent Information Technology Application Workshops, Hangzhou, China.

Zhuang, Z. Y., Churilov, L., Burstein, F., & Sikaris, K. (2009). Combining data mining and case-based reasoning for intelligent decision support for pathology ordering by general practitioners. *European Journal of Operational Research, 195*, 662-675.

Knowledge-Oriented Applications in Data Mining

Edited by Prof. Kimito Funatsu

ISBN 978-953-307-154-1
Hard cover, 442 pages
Publisher InTech
Published online 21, January, 2011
Published in print edition January, 2011

The progress of data mining technology and large public popularity establish a need for a comprehensive text on the subject. The series of books entitled by 'Data Mining' address the need by presenting in-depth description of novel mining algorithms and many useful applications. In addition to understanding each section deeply, the two books present useful hints and strategies to solving problems in the following chapters. The contributing authors have highlighted many future research directions that will foster multi-disciplinary collaborations and hence will lead to significant development in the field of data mining.