



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Development and Test of Fixed Average K-means Base Decision Trees Grouping Method by Improving Decision Tree Clustering Method

¹Jai-Houng Leu, ²Chih-Yao Lo and ²Chi-Hau Liu

¹General Education Committee, Yu-Da College of Business Miao-Li, Taiwan 36143, Republic of China

²Department of Information Management, Yu-Da College of Business Miao-Li, Taiwan 36143, Republic of China

Abstract: New analytical methods and tools which were called FAKDT (Fixed Average K-means base Decision Trees) on human performance have been developed and they make us look at the Enterprise in different aspects in this study. Decision Tree Clustering Method is one of the data mining methods that have been applied widely in different fields to analyze a large amount of data in recent years. Generally speaking, in the human resource incubation of an enterprise, if employees of high learning potential, high stability and high emotional quotient are selected, the return of investment in human resources will be more apparent. If employees of the above mentioned traits can be well utilized and incubated, the industry competitiveness of the enterprise will be enhanced effectively. From the personality specialty point of view, its function is to predict the efficiency of the personal achievement in correlation to his some implying personality specialties (blood group, constellation, etc.). The main purpose of this research is to get the useful information and important message about human performance from their historical records with this method. The Decision Tree Clustering Method data mining skills were improved and applied to get the critical factors that affect the human traits for its feasibility in this study.

Key words: Decision tree, K-means, human traits

INTRODUCTION

Generally speaking, in human resource incubation of an enterprise, if employees of high learning potential, high stability and if employees of the above mentioned traits can be well utilized and incubated, the industry competitiveness of the enterprise will be enhanced effectively. It is hoped that through data mining technique, valuable knowledge or message can be delved from massive historical data. In this study, data mining tool in associated with different thinking is going to be used for the investigation of factors affecting employee's personal job performance.

Data mining means a kind of data analysis method to, from massive, incomplete, hazy and random real applied data, acquire information that people do not know in advance but has its potential meaning so as to achieve effective statistical purpose; since it can be applied in many field such as marketing and sales, stock analysis, customer categorization, space analysis and network document classification, and fruitful results have been obtained, it thus has drawn the attention and research efforts of people from many directions.

In the point of view of organization behavior, the function of human trait is to predict and describe real behavior and performance capability. The first motivation of this study is to find out what kinds of implicit human traits (blood type, constellation, etc.) will affect job performance. The second motivation of this study is to use the clustering method of Fixed average distribution rule to improve the selection of decision tree on node attribute; furthermore, this study is named FAKDT so as to accelerate the efficiency in decision making.

Data mining technique and tool is used to investigate the job performance of the employee of a technology company with stock listed in the stock market. Through the quantitative analysis by rule and the result found by tool mining, the key factors that affect personal performance evaluation is understood; moreover, the index is used to verify if the meaning of improvement exists. Finally, through experiment and analysis, principles that are able to be used to improve employee's performance are summarized and acquired.

Practically, we can see that competence and personality traits are commonly emphasized in human resource selection and qualification process because

these totally different individual's variables can be used as prediction indexes for future job performance. Although personality test is not the best human resource selection tool, but many researchers have found that appropriate collection of personality data will be helpful to the making of selection decision making; in addition, since personality data is not related to other selection tool (for example, recognition competence test), hence, the use of personality test will be able to increase and enhance the prediction power of job performance. However, in recent years, personality test is gradually accepted mainly because the emergence of The Big Five Models have created consensus among personality psychologists of different point of views.

Performance is an evaluation of the degree of achievement of certain goal; in organization behavior (Campbell, 1990), job performance means the entire performance of an individual or organization in three aspects such as Efficiency, Effectiveness and Efficacy.

Factor affecting a person's job performance can be divided into three aspects (Korman, 1977):

- **Job motivation:** Work motivation will affect a person's working attitude, which in turn will affect a person's behavioral performance and job performance
- **Skill:** Whether a person possesses the skill and competence to complete a job will also affect the job performance, which includes the language capability, machinery capability and innovation capability; these different types of capabilities will have different relative importance to different job
- **Role recognition:** That is, a person can precisely feel what is the role needed in the job. Here Following is Research structure in Fig. 1

Model formulation and problem definition: Clustering method (Tseng *et al.*, 2005) is an Unsupervised

Learning method, that is, data won't be distinguished in its category in manual way, or data to be analyzed won't be pre-treated according to its attribute, that is, related clusters of the data will be found out by the system automatically and independently according to data distribution; in data mining application, clustering method is used as a pre-processor for data analysis; that is, for the data to be analyzed, those that have the same attributes and characteristics will be clustered, and through clustering result, the characteristic of each cluster can be effectively highlighted, that is, different cluster will have different characteristic displayed; then according to the designated condition, data is found among certain clusters, and the distribution and trend of all the data can then be seen, which is very helpful to the decision maker to perform data analysis and decision making. In addition, data is clustered according to data attribute so that similar data can be gathered together, hence, in each analysis, only specific cluster needs to be targeted at instead of the consideration of the entire database; for example, the bank can cluster the customers according to their age, income and residence place, etc., by doing so, bank can understand the best customer cluster so as to provide the customer with the most suitable product and service.

The so-called clustering problem is to suppose n original data points $p_i, i = 1 \sim n$ that belongs to d dimensional space (R^d). The main purpose is to divide the original data into k clusters (k is an integer), and the similarity among the members in a cluster is higher than that of members in other cluster; therefore, there are approximate solutions for all kinds of clustering problems currently.

K-means clustering is a very frequent method for solving clustering problem because its concept is easy to understand and easy to be realized. The purpose is to find out points that are similarly on the same d dimensional space (R^d), called center or centroid, and the square

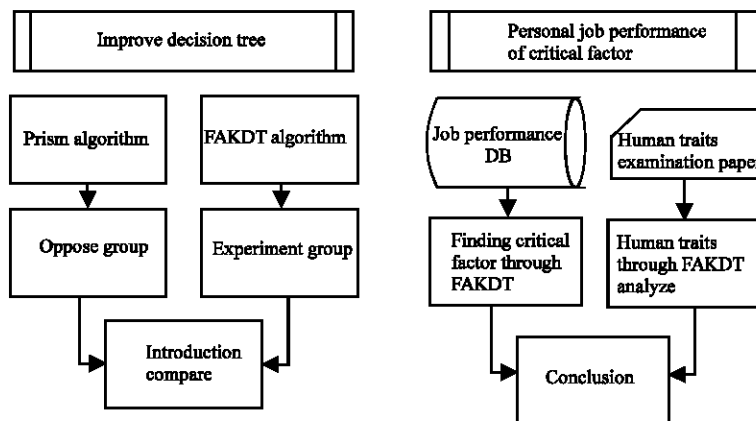


Fig. 1: Research structure diagram

errors, which are also called k-mean errors, from n points (p_i, i = 1 ~ n) to their respective nearest center should be minimized; the formula is as in the following:

$$\sum_{i=1}^n (p_i - c_m)^2, f(i) = 1 \sim k$$

Cf (i) is the center coordinate of the cluster p_i belongs to. The basic algorithm is:

- Decide number k to be clustered (This is assigned manually and is of fixed value)
- Use experience rule or random way to generate k centers (c_j, j = 1 ~ k)
- Distribute n points (p_i, i = 1 ~ n) to the centers that are closest to them respectively
- In each cluster, the coordinates of points belong to it will be averaged to become a new coordinate of that cluster
- Sum up the square of the distance of the point in each cluster to the new center coordinate to get the new k-means error
- Repeat step 3 until certain stop condition is met

The time complication of this method is at step 3 because distance has to be calculated between each point and k clusters so as to find out the cluster that has closest distance to that point; distance has to be calculated in d dimensional space each time, O (nkd) is thus needed. The fourth step 4 is to find the average of the points that belong to each cluster, k clusters will then have n points, hence, it is actually equal to one time of calculation of the original input n points, O(nd) is then needed. Fifth step is to find the distance 5 to the new center, which is very similar to fourth step and O(nd) is then needed; however, how many times step 3, 4 and 5 are to be repeated is not certain but depending on the input data quantity, stop condition, k value, etc. This method is widely used in data mining, image segmentation, photo recognition and artificial intelligence, etc., however, since it is so time-consuming, there are many improvements performed by researchers.

In this study, decision tree based on entropy is taken, which was a decision tree algorithm method ID3 (Iterative Dichotomiser 3, the precursor of C4.5) developed by an early Australian scholar Quinlan and Kohavi (2002). In this method, Information Gain is used as bifurcation principle; but it is found that when it is applied in real case, information gain will prefer to select variable with more selection items, and over-learning is usually the result. In order to improve this system error, (Quinlan and Kohavi, 2002) had re-defined a formula for gain ratio to replace the original bifurcation principle; but no matter it

is what version, the most basic content is to use the so-called entropy concept as the bifurcation principle of decision tree. The gain ratio formula is as in the following:

$$\text{Gain ratio} = \frac{(\text{Entropy before} - \text{Entropy after})}{\text{Split gains}}$$

Information theory: If an event has k results with corresponding probability P_i, then the information quantity acquired I (seen as entropy) after the occurrence of this event is:

$$I = - (p_1 * \log_2 (p) + p_2 * \log_2 (p_2) + \dots + p_k * \log_2 (p_k))$$

Information gain: If the classification mark (Y) can be divided into two types (success and failure), and X is prediction variable (category attribute: k category), n is the total sample number (n₁ is the number that has success mark in the total sample number); after using X variable to classify sample, m_i is the total number of sample in X = i category (m_{i1} is the number in X = i category that has success mark). According to variable X, the information gain to divide n samples into m₁, m₂, ..., m_k is:

$$\text{Gain (x)} = I (n, n_1) - E (x)$$

Where:

$$I (n, n_1) = - ((\frac{n_1}{n}) \log_2 (\frac{n_1}{n}) + (1 - \frac{n_1}{n}) \log_2 (1 - \frac{n_1}{n}))$$

$$E (x) = (\frac{m_1}{n}) * I (m_1, m_{11}) + (\frac{m_2}{n}) * I (m_2, m_{21}) + \dots + (\frac{m_k}{n}) * I (m_k, m_{k1})$$

In this study and in the selection of more influential index, the largest information gain value is used as the first consideration and as the span reference in the judgment range.

And following is sample of personality traits questionnaire of this study.

- What kind of personality trait can be used to describe you?
 - Perfectionist Person who is willing to help others
 - Practitioner Romanticist Knowledge pursuer
 - Peace maker
 - Root cause finder Hedonist Leader

Note: The basic personality traits of category 9 are as in the followings:

Type 1: Perfectionist: Pursue for perfection with principle, not easy to compromise, clearly identify wrong

and right, just, self-controlled, is afraid of mistakes, is picky, has high requirement on himself/herself and others.

Type 2: Person who is willing to help others: Sensibility, willing to help others, friendly, considerate, sympathetic, usually feel not paying enough to others, willing to sacrifice, strong sense to possess.

Type 3: Practitioner: Like to win, ambitious, confident, charming, usually welcome by others, swanky, good leadership, eager for success with strong goal, over-emphasis on the appearance, irritability.

Type 4: Romanticist: Pursue for romanticism, like to imagine, passionate, sensitive from the bottom of the heart, introversive, sensitive, emotional, easily get depressed, easily get into dispirited mood and can not help himself/herself.

Type 5: Knowledge pursuer: Perfectionist, enthusiastic in knowledge pursuit, bravery for reform, strong innovation, good insight, self-isolation, too caring on details, like to dispute, lack of mobility.

Type 6: Root cause finder: Responsible, reliable, calm, insecure feeling, recognition of the authority, fear of making mistakes, sober and kind, over-cautious, hesitating.

Type 7: Hedonist: Materialist, passionate, finding for exciting things, fear of oppression, pursuit of happiness and satisfaction, versatility, willing to explore new things, flippancy, non-obedient.

Type 8: Leader: Liberalist, willing to take risk, anti-society personality, leader of life, like to combat, confident, delicate mind and thinking, ambition to dominate, trustable.

Type 9: Peace maker: Fear of conflict, conservative, strong self-controlled nature polite, passive, lazy, get adrift.

- What kind of advantages as in the followings do you possess?
 - Strong right or wrong sense
 - Sympathetic
 - Excellent performance
 - Strong sense on everything
 - No fear of reform
 - Responsible
 - Passionate
 - Confident
 - Satisfied
- What kind of disadvantages as in the followings do you possess?

- Pursuit of perfection
- Strong desire to dominate
- Pragmatist
- Like to stay alone
- Stubborn on good things
- Too cautious
- Honest in talking
- Strong ambition to win
- Along with the current situation

- What in the followings are things that you are afraid to face?
 - Make mistake
 - Feel not so needed
 - No feeling of achievement
 - Direction loss
 - Feel helpless
 - Feel isolated
 - Fear of oppression
 - Fear of conflict
 - Fear of becoming the weak side
- What kinds of the followings are you dreaming of owning them?
 - To be a perfect person
 - Desire of being relied on
 - Desire of getting accepted
 - Desire of feeling secure
 - To pursue happiness and satisfaction
 - Desire to dominate
 - Search for self-recognition value
 - Desire of owing professional knowledge
 - Pursue of peacefulness

Algorithm formulation: In the study on the improvement of network data mining and clustering technique an effective, simple and improved K-means algorithm has been proposed, and this study is going to base on this basis and use it as a method for attribute selection in decision tree so as to reduce the frequency of pre-treatment in the generation of tree structure.

Fixed average distribution method

- Set up how many clusters are to be clustered, K: No. of cluster.
- Calculate number of data in each cluster (T):

$$T = D \text{ (Total number)}/K$$

- In the applied matrix, take the central value of each cluster:
 - (No. of each cluster * (N-1) + 1 ~ Number of each cluster * (N)
- Perform again the clustering until no clustering is needed.

For each attribute, one S value is calculated to evaluate the spacing distance of that attribute in the clusters; when

the value becomes larger, it means that the attribute has better segmentation result, by doing so, the attribute with largest S value is selected as the node attribute.

$$S_j = |c_j^1 - c_j^2|, j = 1, 2, \dots, 1$$

COMPUTATION RESULTS

The verification of FAKDT: <http://www.ics.uci.edu/~mllearn/MLRepository.html> Four files are acquired from the above web address:

- **Abalone:** Through the body feature and attribute of abalone, age can be predicted; there is a total of 4177 data, which includes 8 attributes
- **Auto Mpg:** Through automobile feature and attribute, the gas consumption of automobile can be predicted; there is a total of 398 data, which includes 7 attributes
- **Breast cancer:** Through the attribute converted from breast cell block image, the time for the re-activation of cancer cell can be predicted; there is a total of 194 data, which includes 32 attributes
- **Automobile price:** Automobile price can be predicted through automobile information; there is a total of 205 data, which includes 25 attributes

Generally speaking, for most of the data mining problems, it is preferred to use less regulations or rules to achieve good prediction or classification; to a model tree, a leaf node represents a judgment rule, and the larger the number of lead nodes, the larger the tree structure of the model tree or the larger the number of existed rules; therefore, in this study, the number of leaf node is used to evaluate the size of the tree structure; Table 1 is the test result for the number of lead nodes before the data file trimming of FAKDT and PRISM.

Use FAKDT to find out the keys of employee's performance: First, the basic employee's data and annual performance evaluation data of one technology company is taken and used, then data pre-treatment and simplification is applied and employee's performance and attribute data as shown in Table 2 below can be obtained, which includes work number, blood type, constellation, marriage, age and sex. Constellation and age are estimated from date of birth; meanwhile, constellation, from Aquarius to Capricorn, is represented by 1 to 12. For marriage, 1 is married, 0 is unmarried, for sex, M is for male and F is for female.

Among them, age can be appropriately clustered through fixed average distribution method into four groups of group A (<= 30), group B (31~42), group C (43-50) and group D (>= 51). By doing so, the branching-stop condition can be precisely found at the data pre-

Table 1: Comparison of algorithm

Data file	Number	No. of attribute	Prism node No.	FAKDT note No.
Abalone	4177	8	76.7	65.2
Auto Mpg	398	7	38.4	27.8
Breast cancer	194	32	26.2	13.2
Car price	159	25	13.3	9.0

Table 2: Employee's performance attribute table

Emp No.	Sex	Blood	Age	Astro	Married	Score
a00828	F	A	A	12	1	86
a00829	F	O	C	4	0	88
a00830	M	O	A	3	0	75
a00831	M		A	4	1	80
a00832	M	B	C	7	0	69
a00833	M	O	D	9	1	77
a00835	M	A	A	1	0	75
a00836	M	O	B	7	0	76
a00837	F		C	8	1	79
a00838	F	AB	B	8	1	80
a00839	F	B	B	9	0	78
a00840	F	O	D	3	1	77

Table 3: Attribute and its information profit-gaining value

Sr. No.	Astro	Blood (value)	Married (value)	Age (value)	Sex (value)
1	0.668	A (0.623)	1 (0.541)	A (0.251)	M (0.231)
2	0.754	B (0.415)	0 (0.357)	B (0.462)	F (0.124)
3	0.882	O (0.752)		C (0.651)	
4	0.452	AB (0.456)		D (0.546)	
5	0.578				
6	0.364				
7	0.641				
8	0.849				
9	0.225				
10	0.345				
11	0.584				
12	0.182				

Table 4: Constellation distribution table

Signs	Frequency	Percentage
Aries	16	5.0
Taurus	28	8.8
Gemini	23	7.3
Cancer	25	7.9
Leo	26	8.2
Virgo	41	12.9
Libra	29	9.1
Scorpio	25	7.9
Sagittarius	35	11.0
Capricorn	27	8.5
Aquarius	17	5.4
Pisces	25	7.9
Total	317	100.0

Table 5: Personality traits distribution table

Personality traits	Frequency	Percentage
Perfectionist	50	15.8
Willing to help others	46	13.6
Practitioner	31	9.8
Romanticist	34	10.7
Knowledge pursuer	30	9.5
Root cause finder	31	9.8
Hedonist	54	17.0
Leader	18	5.7
Peace maker	26	8.2
Total	317	100.0

treatment stage, which saves a lot of time in the repeated pre-treatment.

Table 6: Nine personality traits and twelve constellation distribution

	Capricorn	Aquarius	Pisces	Aries	Taurus	Gemini	Cancer	Leo	Virgo	Libra	Scorpio	Sagittarius
Type1	5.000	5.000	5.000	0.000	7.000	2.000	0.000	5.000	9.000	4.000	2.000	6.000
50	10.000	10.000	10.000	0.000	14.000	4.000	0.000	10.000	18.000	8.000	4.000	12.000
Type 2	1.000	2.000	2.000	3.000	2.000	1.000	4.000	4.000	9.000	3.000	8.000	4.000
43	2.325	4.651	4.651	6.977	4.651	2.326	9.302	9.302	20.930	6.977	18.605	9.302
Type 3	4.000	0.000	2.000	4.000	4.000	1.000	3.000	2.000	6.000	4.000	1.000	0.000
31	12.903	0.000	6.452	12.903	12.903	3.226	9.677	6.452	19.360	12.903	3.226	0.000
Type 4	1.000	1.000	7.000	2.000	2.000	3.000	5.000	0.000	1.000	7.000	4.000	1.000
34	2.941	2.941	20.588	5.882	5.882	8.324	14.706	0.000	2.941	20.588	11.765	2.941
Type 5	6.000	0.000	0.000	0.000	4.000	0.000	3.000	0.000	5.000	5.000	3.000	4.000
30	20.000	0.000	0.000	0.000	13.333	0.000	10.000	0.000	16.670	16.667	10.000	13.333
Type 6	3.000	2.000	1.000	0.000	4.000	3.000	5.000	2.000	5.000	1.000	0.000	5.000
31	9.678	9.452	3.226	0.000	12.993	9.677	16.129	9.452	16.130	3.226	0.000	16.129
Type 7	4.000	2.000	6.000	4.000	0.000	7.000	5.000	8.000	3.000	0.000	5.000	10.000
54	7.407	3.704	11.111	7.407	0.000	12.963	9.259	14.815	5.556	0.000	9.259	18.519
Type 8	3.000	3.000	0.000	3.000	3.000	0.000	0.000	3.000	0.000	0.000	0.000	3.000
18	16.667	16.667	0.000	13.667	16.667	0.000	0.000	16.667	0.000	0.000	0.000	16.667
Type 9	0.000	2.000	2.000	0.000	2.000	0.000	0.000	2.000	3.000	5.000	2.000	2.000
26	0.000	7.692	7.692	0.000	7.692	23.077	0.000	7.692	11.360	19.231	7.692	7.692

We use FAKDT program to continue the classification operation with data source an employee's performance file as in Table 2 and the result is a text file output generated from the attribute value of high and low profit-gaining value, which is summarized and as shown in the following Table 3.

Questionnaire analysis for personality traits: In this study, a total of 350 survey questionnaires have been issued with effective questionnaire of 317 copies and ineffective questionnaire of 33 copies, with a response rate of 90.57%. Meanwhile, the basic data and personality traits of the effective questionnaire sample will be analyzed.

General personality trait classification is done through, for example, blood type, constellation, marriage, sex or even Chinese Zodiacs; moreover, some use combination of these factors together, for example, if blood type is mixed with constellation, 12×4 changes will be generated; therefore, in this study and in the personality trait analysis part, the personality trait of constellation will be analyzed first, and then it will be compared to the result using data mining technique. We statistics constellation distribution in Table 4. And personality traits distribution by questionnaire in Table 5 and personality traits crossing twelve constellation distribution in Table 6.

CONCLUSIONS

This research the personality special characteristic discrimination is nine kinds, the union constellation special characteristic analyzes one by one, develops various personalities special characteristic to be partial in the following constellation, and aims in the table to list the constellation, corresponds various constellations the staff

Table 7: The human characteristic by FAKDT mining and result of the personality traits by questionnaire compares

Human traits	Constellation	Average of job performance	Sort by job performance	Sort by FAKDT mining
Perfectionist	Virgo	86.3	1	1
Willing to help others	Scorpio	79.5	4	4
Practitioner	Taurus	82.1	3	2
Romanticist	Taurus	75.3	6	8
Knowledge pursuer	Capricorn	74.2	7	9
Root cause finder	Cancer	83.2	2	3
Hedonist	Sagittarius	72.1	8	5
Leader	Leo	77.6	5	7
Peace maker	Gemini	69.8	9	6

examination results average score and in the material exploration the constellation attribute weight level in Table 7.

Judge on constellation according to its traits and then compare it with nine personality traits and summarize it, we will get the following analysis result:

Perfectionist: The representative constellation Virgo, Taurus. The world can be clearly divided into black and white. Right is right, wrong is wrong. You must be filled with justice in the mind and be well self-controlled; you must do things efficiently.

Person who is willing to help others: The representative constellation: Virgo, Scorpio. Sensibility, enthusiasm, friendliness, make-people-happy, usually feel that he/she pays not enough to others, willing to help others, willing to sacrifice, strong desire to possess.

Practitioner the representative constellation Taurus, Libra. Emphasize on fame and profit, pragmatist and care about his/her own performance in front of others; like people to see the best side of him/her.

Romanticist: The representative constellation: Libra, Cancer, Pisces. Romantic, filled with imagination, like to express personal feeling through beautiful things; introversive and emotional, easy to get hesitating and indulgent, pursue for unique experience

Knowledge pursuer: The representative constellation: Taurus, Libra and Capricorn. Enthusiastic in pursuing knowledge, like to analyze things and investigate abstract concept so as to build ideal structure.

Root cause finder: The representative constellation: Leo, Scorpio. Recognize and obey the authority, responsible; but when opponents are met, this kind of person might easily get into a contradiction of over-bearing and attacking; hence, this type of person might become hesitating and over-cautious.

Hedonist: The representative constellation: Leo, Capricorn and Sagittarius. Extroversive, non-oppressive, person with wideview, materialist, like to explore new and fresh things, know very well self-amusement.

Leader: The representative constellation: Leo, Gemini. Totally liberalist, willing to take risk, a person that is good at steering, self-will, like to combat, will not obey under the authority but will build a new kingdom.

Peace maker: The representative constellation: Leo, Aquarius. Willing to adapt to the real situation without finding a change, very passive; not very enthusiastic about life, with strong faith in fatalism, hence, this type of person will resign himself/herself to his/her fate; this type of person emphasizes on the advantages of other's situation but can not handle well on the problems of people around him/her or face himself/herself faithfully.

In the selection and improvement of decision tree algorithm aspect, although clustering method has the advantage of fast execution speed, yet there are many researches show that clustering method is less stable, for example, it can not get convergent; therefore, in the literature of recent years, there are many researches that add other concepts into clustering method. In traditional decision tree, if the classification result is not satisfied, pre-treatment stage must be gone back to for the re-adjustment of the pre-treatment method and for the re-performing of data mining job, and the process must be repeated until satisfaction has been reached. However, this method is very tedious and resource-consuming; in FAKDT method, fixed average clustering method is used

first for the attribute sorting and clustering, then general decision tree operation is performed; the tree structure used in FAKDT will be smaller.

In the verification of FAKDT and in certain data files, this study still can not find out the obvious advantageous and disadvantageous cause of FAKDT, maybe more file examples need to be adopted; therefore, in the subsequent research, other information can be used as proof, for example, data characteristics and the relationship among attributes, etc. In the comparison of data mining and personality trait, only constellation is aimed at for the analysis and comparison, hence, in the subsequent job, other factors such as blood type, Zodiacs, etc., can be added for more complicated cross-comparison, maybe there will be more interesting findings found.

In the comparison of data mining and personality traits survey and analysis and in the constellation personality trait aspect, the result displayed by data mining technique is consistent with the result summarized from questionnaire survey; moreover, general people's perception on constellation trait is also met, for example, most of Taurus who pursue for perfection and emphasize on fulfilling power have very high job performance. This system is designed in user friendly interface and easy operation orientation, the user only needs to feed the clustering data into the system, and then the needed clustering data can be obtained in the shortest time.

In the future, this system will be able to be applicable to aspects such as consumer habit, merchandise classification, customer clustering and stock recommendation so that the largest effectiveness can be exploited.

REFERENCES

- Campbell, J.P., 1990. Modeling the Performance Prediction Problem in Industrial and Organizational Psychology, Handbook of Industrial and organizational Psychology. 1st Edn., Consulting Psychologists Press, Palo Alto, CA, pp: 687-732.
- Korman, A., 1977. Organization Behavior, Englewood Cliffs. 1st Edn., Prentice Hall, Inc., New Jersey.
- Quinlan, J.R. and R. Kohavi, 2002. Data Mining Tasks and Methods: Classification: Decision-Tree Discovery. 1st Edn., Oxford University Press Inc., New York, USA., ISBN: 0-19-511831-6, pp: 267-276.
- Tseng, H.H., H.M. Tsai, T.H. Su and C.J. Tseng, 2005. Data Mining. 1st Edn., Flag Publishing Co. Ltd., Taipei, ISBN: 9574422364, pp: 122-143.