Analysis of Abnormality Diagnosis in Emergency Medicine by Integrating K-means and Decision Trees-a Case Study of Dongyang People's Hospital in China

Zhong Lv
Dongyang Hospital of Wenzhou Medical College, Dongyang, China

Jinkan Du
Dongyang Hospital of Wenzhou Medical College, Dongyang, China

Wen-Tsann Lin
National Chin-Yi University of Technology/Graduate Institute of Industrial Engineering & Management, Taiwan, R.O.C

Shen-Tsu Wang
National Pingtung Institute of Commerce /Department of Commerce Automation and Management, Taiwan, R.O.C

Chia-Ching Chang
National Chin-Yi University of Technology/Graduate Institute of Industrial Engineering & Management, Taiwan, R.O.C

Meng-Hua Li
Taiwan Shoufu University/Department of Industrial Engineering and Management, Taiwan, R.O.C

Zheng-Han Zhuang
National Pingtung Institute of Commerce /Department of Commerce Automation and Management, Taiwan, R.O.C

***Abstract*—The performance of a triage system can facilitate patient classification in an emergency department, enabling patients in critical condition to receive better medical care; therefore, more perfect allocation and use of resources of emergency medical treatment are required. The correctness of nurses and doctors is related to triage medical care quality, patient satisfaction, and life safety. Hence, how to effectively extract experience by data mining and triage in the background of continuously increasing numbers of emergency patients is an issue worth exploring. Based on the case of Dongyang People's Hospital in China, this study established a triage prediction model from process construction, parameter selection, and sampling, and randomly generated 501 samples of patients from the emergency database for cluster analysis (Ward's method and K-means) and decision trees analysis upon data mining. The findings of this study show that the triage categorization of nurses is higher than that of doctors and most abnormal diagnoses occur to patients not examined on the date of admittance. The vital signs of pulse and temperature are more discerning. According to the confidence and support proportion, this study proposed seven association rules.**

***Index Terms*—emergency department; triage medical care quality; data mining; K-means; decision tree**

Corresponding author: Shen-Tsu Wan,
d91782@oz.nthu.edu.tw

I. Introduction

The emergency department is always the forefront in the treatment of patients, as it provides 24x7 professional emergency medical services to realize the emergency treatment of the patients for the first time. The department consists of doctors, nurses, various technicians, social workers, ambulance technicians, administrative staff, workers, and voluntary workers. Regardless of any difficulties, they will come forward for patient emergency. Indwelling observations or surgery may be conducted at any time in the department, which is like a small hospital within the larger one.

The emergency department visit procedure is, as follows: "when a patient is admitted to the emergency room, the patient should be immediately categorized in accordance with the triage system ("triage", as the name suggests, means that the emergency treatment professionals determine the emergency priorities and

doi:10.4304/jsw.9.10.2764-2770

sequence of treatment of the emergency patients according to the causes, patient c/o, medical history, disease severity, and urgency. Various important patterns and indicators are used as benchmarks for classification, and subsequent large-scale disaster treatments and hospital classifications of emergency treatment patients are attributed to the triage system.) [2]. The number of emergency medical treatment patients has continuously increased over the years, and reached 20,544 people/visits, an increase of 72% as compared to 11,916 people/visits in 1994. The dramatic increase in patients result in chaotic, busy, and packed situations of the emergency departments, as patients do not know the medical diagnostic process or the medical system of classification by departments.

The triage categorization system, in accordance with the hospital emergency department assessment standards of the Ministry of Health and Welfare[3] , classifies emergency treatment patients into five categories: Category I (Should be treated immediately), Category II (Should be treated within 10 minutes), Category III (Emergency should be treated within 30 minutes), Category IV (Secondary emergency that should be treated within 60 minutes), and triage Category V (Sub-emergency that should be treated within 120 minutes).

In recent years, the popularization of information applications and industrial demand has seen the rapid development of data mining technology. Data mining means the "actual process of discovering non-obvious, unprecedented, and potentially useful information" [4]. Data mining is able to determine the hidden information of data, thus, data mining is actually a part of Knowledge Discovery. Data Mining uses many statistical analysis and modeling methods to determine the useful patterns and relationships of data. Reference [5] is data mining software in the Java language developed by the University of Waikato in New Zealand, which has various tools, including Preprocess, Classify, Clustering, Associate, Select attributes, and Visualization. It can be seen that data mining has been applied in daily life and many scholars have been developing its application in industrial and medical fields. For example, Reference [6] applied the data mining clustering analysis technology in the discussion of the characteristics of revisiting patients within 72 hours after emergency treatment; Weng [7] applied data mining technology in customer relationship management-a case study of children's dental clinics. Reference [19] used a new text clustering algorithm is presented based on K-means and Self-Organizing Model (SOM). Firstly, texts are preprocessed to satisfy succeed process requirement. Secondly, the paper improves selection of initial cluster centers and cluster seed selection methods of K-means to improve the deficiency of K-means algorithm that the K-means algorithm is very sensitive to the initial cluster center and the isolated point text. Thirdly the advantages of k-means and SOM are combined to a new model to cluster text. Reference [20] considered an improved k-means algorithm based on optimized simulated annealing is used to segment the stations of Hangzhou Public Bicycle System. The optimized simulated annealing(SA) algorithm is used to assign k-means initial cluster centers. Practice examples and comparison with the traditional k-means algorithm are made. The results show that the proposed algorithm is efficient and robust. The research result has been applied in Hangzhou.

This study applied the K-means decision tree and data mining technology in the discussion of abnormal triage diagnoses of emergency medical treatment. First, this study collected academic journals. Moreover, we cooperated with Dongyang People's Hospital in China to understand the overview of the implementation of the triage system in the emergency department, and obtained basic statistics and patients' data from the emergency treatment database. Afterwards, through the study of administrative executives, doctors, and nurses of Dongyang People's Hospital emergency department, coupled with the database provided by the hospital, we conducted data mining. By applying data mining technology integrated with theory and practice, we aimed at improving emergency medical treatment, with the purposes of this study summarized, as follows:

1) Beginning with data management, this study applied the system to the emergency department triage prediction model, and specifically described the decision-making processes of nurses.

2) To discuss the relevance between triage and abnormal diagnoses, and analyze the contributions of various triage parameters for the reference of nurses in the implementation of triage.

3) Using Hierarchical clustering (Ward's method) and Partitioning clustering (K-means algorithm) to establish abnormal triage diagnosis clusters and decision trees (C4.5 algorithm) and determine the association rules of the occurrence of abnormal diagnoses.

4) To apply data mining technology to enhance triage consistency, and provide quantified and scientific information for the decision-making process of triage for the reference of subsequent researchers and clinical inspection.

II. Definition Of Triage Process

In this study and upon discussion, the process construction, parameter selection, and sampling were confirmed by literature and approved by the administrative and professional staff of the emergency department of Dongyang People's Hospital. The triage research architecture and research methods are illustrated, as follows.

A. Triage Categorization Prediction Models

The purpose of establishing the emergency department triage system of the triage prediction model is: "no delay in the treatment of patients in actual emergencies due to too many visits, as well as the right person, in the right time, at the right place, and the use of the right resources. The procedure of the emergency department is: "when a patient is admitted to the emergency room, the nursing staff will conduct triage on the patient, according to the degree of critical conditions of the patient, before allocating the patient to different wards (departments). After the end of treatment by the doctor, the patient will be arranged according to the conditions, and the doctor will assess the triage judgment of the nursing staff. " If the nursing staff and the doctor are consistent in triage judgment, the triage is correct. If the nursing staff judges the conditions as being more serious than the determination of the doctor, it is called an overestimation. Otherwise, it is known as the underestimation [8].

The standards of the nursing staff are based on the patient's c/o (complains of), medical history, general appearance, vital signs, symptoms and signs, and physical assessment results. The patient's medical history, general appearance, symptoms and signs, and physical assessment results could not be quantified; therefore, the emergency department integrates them as patient c/o (patient's description of own condition) for the record. The vital signs include respiration, temperature, pulse, diastolic blood pressure (diastolic, dias.), systolic blood pressure (systolic), and SaO2. Regarding respiration, as few patients have respiration abnormalities, once any patient respiration abnormality is identified by the nursing staff, it should be included in the record of c/o, and measured.

During the triage decision-making process, in addition to the influences of the nursing staff and doctor, patient condition is one of the major considerations of the occurrence of abnormal triage diagnosis, for example, patient's condition suddenly worsening, recovery, or worsening after leaving the hospital will have an impact on triage. Therefore, the interval of patient triage is also included as a parameter for analysis in this study; regarding patients after receiving treatment, they may be categorized into patients of emergency pediatrics, emergency internal medicine, emergency rescue, emergency surgery, emergency ophthalmology and otorhinolaryngology, emergency ophthalmology, or emergency orthopedics.

After summarizing of the emergency department processes and parameters, we simplified the process into Input, Process, and Output. First, this study obtained the data of parameters from the database of the emergency department, and discussed with the professionals of the emergency department about the extreme values, and omissions of data, in order to confirm removal or modification before inputting the summarized patient data into the data mining tool for subsequent analysis to enhance the triage decision-making consistency. The triage prediction model is described, as shown in Fig. 1.

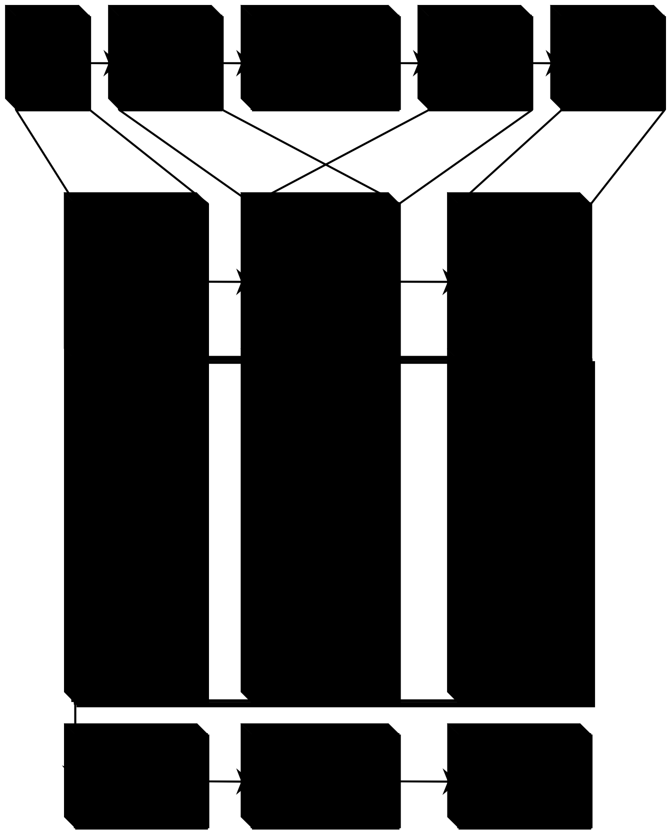

Figure 1. Triage prediction model

B. *Data Mining Prediction Model*

Data mining is actually a process of the discovery of the non-obvious, the unprecedented, and the credible, as well as potentially useful information, such as trends and patterns taken from useful knowledge in order to analyze the data and obtain a more in-depth understanding. It is expected to result in interesting knowledge obtained from the database of random, chaotic, complex, and large amounts of data, including unexplainable or unidentified causal relationships; in general, data mining has two major functions, one is to predict future trends by modeling, and the other is to determine any unknown patterns from the data; data mining models include: Classification, Estimation, Prediction, Affinity grouping, Clustering, Description, and Profiling [9]. This study obtained the data of patients of abnormal triage from the patient treatment database of the emergency department. After the data was corrected, we applied cluster analysis to cluster the patients of abnormal diagnosis of the same nature for analysis, using the classification method to determine any hidden rules. The data mining process in this study is as shown in Fig. 2.

C. *Cluster Analysis*

Clustering analysis is also known as data segmentation or unsupervised classification. It is a multi-variate statistical analysis technology used mainly to cluster data entries or data points, where observations or cases of the data sets are reduced into a few sets in order to realize the higher intra-group similarities of the data points, rather than the inter-group similarities of the data points[10]. There are three main types of clustering analysis methods:

1) Partition clustering is to directly categorize data entries into a few mutually disconnected clusters, in order that the similarities between the records of the cluster and the cluster center are higher than the similarities with other cluster centers.

2) Hierarchical clustering is to combine the smaller clusters of high similarity into greater clusters or vice versa.

3) Two-stage clustering is to integrate the hierarchical and non-hierarchical methods.

This study used the hierarchical Ward's method clustering results as the initial clustering before using the K-means algorithm of partitioning clustering for the adjustment of clusters. We applied SPSS 10.1 in Cluster analysis, with the two-stage hierarchical clustering, as proposed by Anderberg [11], and illustrated and analyzed, as follows:

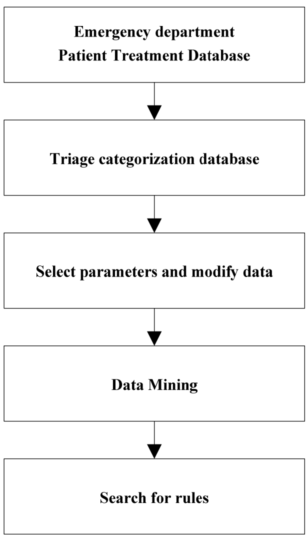

Figure 2. Data mining prediction model

D. K-Means Clustering Analysis

Clustering analysis (Cluster) is the unsupervised machine learning of data mining, and is mainly to separate population samples into homogenous sub-clusters. The ideal results achieve "homogeneity within the cluster and heterogeneity between clusters", that is, the differences of the samples in the same cluster should be as small as possible, while the differences of samples between clusters should be as great as possible [12]. In statistics, the basic principle of the clustering method is to use the similarities as the partition objective, and the clustering methods can be mainly divided into Hierarchical Clustering and Partition Clustering methods. The hierarchical algorithm is not a one-time clustering method. Generally, it uses tree-like structures and the practices include agglomerative and divisive methods. Agglomerative algorithm starts by regarding each object as an independent cluster, and continuously measures and combines similar clusters until all objects are combined into a group or a number of groups that can satisfy the needs of the user to complete the entire clustering. It is a bottom-up clustering method [13]; while the divisive method is exactly the opposite. It uses the top-down method by regarding all the data as one group, and gradually separates the different data [14].

The K-Means algorithm is the most widely known non-hierarchical clustering analysis method. Formally proposed by MacQueen [15], it determines the optimal clustering center by repeating the process until convergence. The computation methods are shown, as follows:

Step1: set a fixed value for K, which represents the number of samples to be clustered.

Step2: from n data points, by random numbers, select the K initial central points as the centers of the initial clusters.

Step3: use Euclidean distance to allocate the remaining data points and categorize the data into the nearest cluster centers.

Step4: re-compute the new centers.

Step5: If K new centers are the same as the original centers, stop clustering; otherwise, return to Step3 and continue clustering.

The so-called clustering is a data mining technology that categorizes the data according to a certain condition. The clustering methods can be categorized into (1) divisive clustering; (2) hierarchical clustering; (3) density clustering; and (4) grid clustering

III. PARAMETER MEASUREMENT AND DATA ANALYSIS

A. Sample Selection

This study obtained the patient data of 2012 from the emergency department, and the triage patient distributions are as shown in Table 1: patients of triage Category I account for 8.01% of the total, patients of Category II account for 41.30% of the total, patients of Category III account for 50.16% of the total, and patients of Category IV account for 0.53% of the total; the proportions of the abnormal decisions of triage by category are shown, as follows: Category I 15 cases, Category II 354 cases, Category III 121 cases, Category IV11 cases, for a total of 501 cases of triage screening changes (overestimation and underestimation).

From the basic statistics, it can be concluded that, Category II and Category III patients account for the greater proportions in terms of number of patients and number of abnormal triage decisions. Abnormality decisions seem to be more frequent in the case of Category IV, followed by Category II, Category III, and Category I. To determine the hidden information rules of the patient data, this study used the categories of triage as the targets, and randomly selected 2000 patient samples of consistent triage decisions. To ensure the pre-event probabilities of the cases in various categories as equal, the limit is 519 of the cases of consistent triage Category IV. We used the data of 501 cases of abnormal triage decisions in 2012 as samples for analysis.

TABLE I.
PATIENT TRIAGE SURVEY

2012	Triage Category I	triage Category II	triage Category III	triage Category IV	Total
Triage correct decisions	8,622	44,576	54,268	519	107,985
Triage abnormality	15	354	121	11	501
Abnormality percentage (%)	0.17%	0.79%	0.22%	1.90%	0.46
Total number of patients	8,688	44,801	54,411	580	108,480
Category (%)	8.01%	41.30%	50.16%	0.53%	100

B. *Parameter Measurement*

Regarding parameters, this study used date of visit, visiting card number, name of clinic area, doctor's work number, department name, and diagnosis results, in order to measure five parameters, with three parameters in the decision-making dimension and situations of the patients after leaving the emergency department for clustering analysis.

C. *Vital Signs*

Five parameters of vital signs are of a ratio scale. According to the provisions of the "Adult Triage Scale" of the Ministry of Health and Welfare [16], the decision-making dimensional parameters include the time interval of doctor treatment and nursing staff decisions. The doctor treatment and nursing staff triage decisions include: Category I: resuscitation, in need of immediate treatment, temperature≧41℃or ≦32℃, GCS 3-8 severe comma, systolic blood pressure >220mmHg, or ≦ 80mmHg, Category II: critical, in need of reassessment within 10 minutes, temperature in the range of 40℃ ~ 39 ℃or in the range of 35℃~ 32℃, systolic blood pressure in the range of 220mmHg ~ 180mmHg, Category III : urgent, in need of reassessment within 30 minutes, GCS 9-13 altered consciousness, and temperature<39 ℃ , Category IV: secondary urgent, in need of reassessment of GCS 14-15 within 60 minutes. Category V: non-urgent, in need of reassessment within 120 minutes, which are ordinal scales. Moreover, according to the clinical diagnosis of the doctor, regarding blood pressure: systolic blood pressure in the range of 140mmHg ~ 160mmHg, and diastolic blood pressure in the range of 90mmHg ~95mmHg are critical values, when systolic blood pressure >160mmHg and diastolic blood pressure >95mmHg, it is known as hypertension, when systolic blood pressure <90mmHg ~ 100mmHg, it is known as hypotension. Regarding pulse and concerning adults, when the pulse per minute is above 100, it is called tachycardia. When pulse per minute is below 60, it is known as bradycardia. Both tachycardia and bradycardia can affect the per minute blood volume of the heart. Regarding SaO2, the normal concentration of SaO2 is in the range of 92% to 99%, and too low a concentration of SaO2 will lead to loss of consciousness. By the decision making time interval, it can be divided into the current date triage and non-current date triage, as measured by the nominal scale. Regarding the situations of patients leaving the emergency department, they can be categorized into patients of emergency treatment of pediatrics, emergency treatment of internal medicine, emergency treatment of rescue, emergency treatment of surgery, emergency treatment of ophthalmology and otorhinolaryngology, emergency treatment of ophthalmology, and emergency treatment of orthopedics, as measured by the nominal scale.

IV. DISCUSSION

Different cultures must be considered. When adopting western medical management theories, Chinese medical practitioners may be faced with problems of cultural conflict if no consideration is given to the differences between Chinese culture and western culture. Therefore, hospital managers should introduce Chinese traditional culture as a reference framework, and integrate it with the modern medicine management theories based on the organizational situations in order to display the great effect of applying the traditional culture of China in contemporary matters [17].

The decision tree algorithm is shown, as follows: if there are n types of classification results, the occurrence probabilities are P (X1) … P (Xn), respectively, then the amount of information after the occurrence of the event is:

$$I[P(X_1),\ldots,P(X_n)] = \sum_{i=1}^{n} P(X_i)\log_2 P(X_i) \quad (1)$$

If the classification indicators include P1 … Pn, A represents an attribute; V represents the sample set before attribute testing, V1 … Vk represent the sample set after attribute testing; X1 represents the number of P1 in V …; Xn represents the number of Pn in V; then the information gains of dividing V into V1 … Vk in accordance with the attribute values of A are:

$$\mathrm{Gain}(V) = I(P_1,\cdots,P_j) - E(V) \quad (2)$$

After clustering, the sample points of the same cluster have relatively similar characteristics of abnormal diagnosis; therefore, this study conducted individual analysis of four clusters of samples. Regarding the parameters, this study input the six parameters of vital signs as the decision making nodes. Regarding operations, this study applied Ctree software's C4.5 algorithm in the construction of the decision trees by using four parameters including, Minimum node size, Maximum depth, Minimum support, and Minimum confidence [18].

This study selected nodes of higher proportions of Confidence and Support for discussions with the administrative heads of the emergency department, and found that clusters include patients of triage Category I to Category IV. As a result, the support proportions of nodes are relatively low. However, the administrative heads of the emergency department argued that, the sample points of abnormal diagnosis include patients of triage Category I to Category IV. Hence, mainly based on the confidence proportion and support proportion as the supplement, this study selected seven rules: for Cluster 1 patients, low temperature or high systolic blood pressure may easily result in abnormal diagnosis, for Cluster 2 patients, low pulse may easily result in abnormal diagnosis, for Cluster 3 patients, low systolic blood pressure and diastolic blood pressure may easily lead to abnormal diagnosis; for Cluster 4 patients, high pulse and temperature can easily lead to abnormal diagnosis.

When applying the weights in the decision trees, we discovered an issue: as mentioned in Section 3.1, abnormal triage decisions are more frequent in cases of patients of Category IV and Category V. However, as the number of patients varied from Category I to Category V, the final number of patients of abnormal triage was 0.46%, as determined by applying the weights, which may lead to the final results of the decision trees. However, it can be found from the triage rules, as established by decision trees, that abnormal decisions mainly occur at the critical values between the categories of triage. Such situations are more frequent in the case of parameters of temperature, systolic blood pressure, and pulse. Finally, the abnormality value of patients of triage Category I is 0.17%, the abnormality value of patients of triage Category II is 0.79%, the abnormality value of patients of triage Category III is 0.22%, and the abnormality value of patients of triage Category IV is 1.90%. By applying the optimal abnormality proportions into the number of patients of various categories of triage, the number of patients of triage abnormality is 54,492.65. Whether the abnormal decisions of Category IV and Category V of triage may be caused by fewer patients or higher classification by the nursing staff of the emergency treatment patients is worthy of future exploration.

V. Summary And Conclusions

A. Summary

This study cooperated with the emergency department of Dongyang People’s Hospital in China in order to discuss triage abnormality diagnosis rules. From process construction, parameter selection, and sampling, we constructed an abnormal diagnosis model to review the database, and generated the necessary patient data from the model. Afterwards, we conducted two-stage cluster analysis (Ward’s method and K-means) of 501 samples of abnormal diagnosis from the registration inquiry database, triage database, and medical order inquiry database.

The findings of this study show that the nursing staff usually classifies patients with a higher category, as compared to the doctor. The vital signs of pulse and temperature have greater distinguishing power. Moreover, the non-current date triage of Cluster 2, Cluster 3, and Cluster 4 is more serious. Therefore, patients admitted multiple times to the emergency room, or patients of recurrence, are more susceptible to abnormal triage diagnosis.

B. Conclusions

Based on the data of the patients of the emergency room of Dongyang People’s Hospital in China, we applied data mining technology and proposed a research framework for the reference of scholars. Therefore, to understand the general rules for Taiwan and China regarding emergency department triage abnormal diagnosis, we must collect more samples for analysis by more rigorous methods; moreover, after triage categorization of patients pending emergency treatment, in addition to vital signs, the impact of patient c/o, including patient medical history, general appearance, symptoms and signs, and physical assessment results should be considered. Hence, we also suggest that scholars should consider c/o in future studies.

Finally, we proposed a suggestion to the emergency department: “when a patient is admitted for triage, if the patient vital signs are: temperature above 40.7℃ or below 35.6℃, pulse per minute above or equal to 105 times or below 52 times, systolic blood pressure above 210mmHg or below 68mmHg, and diastolic blood pressure below 90mmHg, then the triage classification should be rigorously assessed according to other parameters, including patient c/o and medical history.”

Acknowledgements

The authors are grateful to Dongyang People’s Hospital in China for providing emergency department-related information, and consent for disclosing the name of the hospital and relevant data, in order to facilitate the completion of this research and the writing of this article.

References

[1] G. Karegowda and A. Siddaganga. “GA based dimension reduction for enhancing performance of k-means and fuzzy k-means: A case study for categorization of medical dataset dvances”. Intelligent Systems and Computing, vol. 201, pp. 169-180, 2013.

[2] W. L. Wang, W. T. Lin, C. H. Lai and H. M. Chen. “Application of data mining technology to enhance emergency medical triage of consistency - a medical center in Taiwan Department of Emergency Medicine Case”. Journal of Quality, vol. 15, pp. 283-291, 2008.

[3] Ministry of Health and Welfare, Executive Yuan, Taiwan, ROC. Hospital emergency department evaluation criteria, 2012.

[4] W. J. Frawley, G. Paitetsky-Shapiro and C. J. Matheus. “Knowledge Discovery in Databases: An Overview Knowledge Discovery in Database”. California, AAAI/MIT Press, pp. 1-30, 1991.

[5] M. J. Berry and G. Linoff. “Data Mining Techniques for Marketing, Sales and Customer Support”. New York: John Wiley & Sons, 1997.

[6] H. C. Huang. “Cluster analysis using data mining techniques to explore branches emergency 72 hours and then return to clinic patients characteristics”. Cheng Vhing Medical Journal, vol. 5, no. 3, pp. 13-20, 2009.

[7] W. Z. Qian. “Application of data mining techniques in

customer relationship management - for example children dentist". Changhua University of Business Administration Thesis, Taiwan, ROC, 2011.
[8] W. T. Lin, S. T. Wang, T. C. Chiang, Y. X. Shi, W. Y. Chen and H. M. Chen. "Abnormal Diagnosis of Emergency Department Triage Explored with Data Mining Technology: An Emergency Department at a Medical Center in Taiwan taken as an Example". Expert Systems with Applications, vol. 37, no. 4, pp. 2733-2741, 2010.
[9] J. Y. Wu and C. H. Chou. "The Study of the Six Major Causes of Death Chronic Diseases by Data mining". J Inf Manag, vol. 18, pp. 187-211, 2011.
[10] HH. Tseng. "Data Mining". Taipei: Flags Press, 2006.
[11] M. R. Anderberg. "Cluster analysis for applications. Probability and Mathematical Statistics". New York: Academic Press, 1973.
[12] Li Zi Instrument. "Hospital outpatient data mining - in Huwei Joseph's Hospital, for example." Nanhua University, Institute of Information Management thesis, Taiwan, ROC, 2000
[13] X. Z. Wu and S. Z. Lai (trans.) "Data Mining data mining theory and practice of customer relationship management skills and science". Taipei: Vico Books Limited, 2001.
[14] C. C. Wang, M. S. Chen and C. J. Hsu. "Consumer Relationship Management Using Cluster Analysis of High Level Health Examination Customer Groups as a Data Mining Technique-Example of a Single Medical Center in Taipei." J Health Manag, vol. 6, no. 1, pp. 49-60, 2008.
[15] J. B. MacQueen. "Some methods for classification and analysis of multivariate observations." Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, pp. 281-297, 1967.
[16] Ministry of Health and Welfare, Executive Yuan, Taiwan, ROC. General Medical Services trips over the year, 2012.
[17] H. N. Yang. "Confucianism and Legalism view of human resource management implications." Journal of Chienkuo Technology University, vol. 28, no. 4, pp. 65-84, 2009.
[18] S. H. Liao. "Knowledge Management Technologies and Applications-Literature Review from 1995-2002." Expert Systems with Applications, vol. 25, pp. 155-164, 2003.
[19] Li, X. "A New Text Clustering Algorithm Based on Improved K_means." Journal of Software, vol.7, no. 1, pp. 95-101, 2012.
[20] H. Xu, J. Ying, F. Lin and Y. Yuan. "Station Segmentation with an Improved K-Means Algorithm for Hangzhou Public Bicycle System." Journal of Software, vol. 8, no. 9, pp. 2289-2296, 2013.