World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:3, No:5, 2009

A Decision Support System for Predicting Hospitalization of Hemodialysis Patients

Jinn-Yi Yeh and Tai-Hsi Wu

Open Science Index, Economics and Management Engineering Vol:3, No:5, 2009 publications.waset.org/3884/pdf

***Abstract*—**Hemodialysis patients might suffer from unhealthy care behaviors or long-term dialysis treatments. Ultimately they need to be hospitalized. If the hospitalization rate of a hemodialysis center is high, its quality of service would be low. Therefore, how to decrease hospitalization rate is a crucial problem for health care. In this study we combined temporal abstraction with data mining techniques for analyzing the dialysis patients' biochemical data to develop a decision support system. The mined temporal patterns are helpful for clinicians to predict hospitalization of hemodialysis patients and to suggest them some treatments immediately to avoid hospitalization.

***Keywords*—**Hemodialysis, Temporal abstract, Data mining, Healthcare quality.

I. Introduction

END stage renal disease (ESRD) is commonly known as uremia. Patients suffer from this disease need to do periodical hemodialysis (HD) all their life to keep alive and healthy. During the long-term dialysis treatment, they are likely to get hospitalization due to careless in daily life or other disease infection. This is also the main reason for unreduced hospitalization rate of hemodialysis patients in the past years. For hemodialysis institutions, a high hospitalization rate means low health care quality of that dialysis center. Therefore, dialysis centers are also thinking about how to reduce the hospitalization rate effectively to improve the service quality. However, the hospitalization rate of hemodialysis patients had no obvious decline in the past years, from the view of preventive medicine, how to prevent and reduce the hospitalization rate of dialysis patients effectively is very important. In this paper we are interested in development of decision support system to predict hospitalization of hemodialysis patients which is delivered by a hemodialysis center on the basis of the process data routinely collected during hemodialysis sessions.

For patients receiving long-term hemodialysis treatment, hemodialysis centers would examine their biochemical data monthly such as hematocrit (Hct), blood urea nitrogen (BUN), and creatinine, and provide professional persons' records and evaluate patients' physiological state and dialysis quality [1]. For hemodialysis time series data, the temporal abstraction (TA) method brought forward by Shahar [2] can integrate with professional knowledge for the process of data analysis. It is especially applicable to high dimensional and massive time series data [3], [4]. Bellazzi et al. [5] successfully combined temporal abstract and association rule algorithm, applied to evaluation of dialysis service quality and execution of temporal data mining. They found out the temporal pattern resulting in unsound dialysis quality effectively. This approach proved quite applicable to clinical time series data. Therefore, we cascaded temporal abstract and data mining techniques in this study to find out the temporal patterns resulting in hospitalization of dialysis patients.

Data mining techniques have become mature hitherto since the beginning in 1990's. The purpose of data mining is to find out potentially useful, newfangled and easily comprehended knowledge from data warehouse. An effective model is then established for predicting, analyzing, and supporting for a decision [6]. The classification method of data mining is a mode of machine learning which is widely applied to clinical medicine region at present. Decision tree is mainly to deduce a prediction model which is applicable to special patient, to predict the result, and to give a support for making medical decisions [7]. The classification method was widely applied to medical data analysis in the last few years, such as forecasting the survival time of renal dialysis patients, extracting features of trauma patients' data, and constructing a prediction model, integrating data mining and case-based reasoning into prognosis and diagnosis of chronic diseases, classifying rules applied to knowledge discovery of data set of cardiovascular diseases [8]-[11]. From these studies, the result of applying data mining classification method to medical field is helpful for decision making. At the same time this is effective to improve future medical techniques and quality services.

This study integrates temporal abstraction method with data mining techniques to analyze biochemical test data of hemodialysis patients. We explore the hospitalization pattern hidden in data, and set up a decision support system to provide professional medical care personnel with rule retrieve. Then the clinicians would be able to judge the probability of patients' hospitalization timely and adopt preventive medical measures to decrease the incidence of hospitalization. Through this system, we hope to alleviate patients' physical and mental burdens and to improve the quality of service of dialysis centers. The rest of this paper is arranged as follows: Material

Jinn-Yi Yeh is with Department of Management Information Systems, National Chiayi University, 580 Sinmin Rd. Chiayi, 600 Taiwan (corresponding author phone: +886-5-273-2899; fax: +886-5-273-2893; e-mail: jyeh@ mail.ncyu.edu.tw).

Tai-His Wu is with Department of Business Administration, National Taipei University, 151, University Rd., San Shia, Taipei, 237 Taiwan (e-mail: taiwu@mail.ntpu.edu.tw).

Open Science Index, Economics and Management Engineering Vol:3, No:5, 2009 publications.waset.org/3884/pdf

and Methods Section describes the hemodialysis and quality of its center, the development of decision support sysytem used in this paper is demonstrated in the Developing Decision Support System Section, Experimental Results Section illustrates the experimental results using the combined approach for hemodialysis patients' data analysis, and conclusion is in the Conclusion Section.

II. Material and Methods

A. Hemodialysis

Treatments of ESRD patients can be classified as hemodialysis and peritoneal dialysis. It is more than 90% of total patients taking hemodialysis session in Taiwan recently. They need to undergo blood depuration (hemodialysis) through an extra-corporal circuit three times a week for four hours. The data accumulated over time for each patient contain the set of variables that are monitored during each dialysis session. Kusiak et al. [8] combined C4.5 decision tree with rough set theory to explore biochemical test data of hemodialysis patients. The data is classified as eight training sets of different types corresponding to eight classsifications for setting weight values of two algorithms. Then the data would be transferred to 16 classsifications. Finally a voting scheme would be used to predict whether the patient could survive for more than 3 years or not.

In order to improve public's health quality, the National Health Insurance Bureau (NHIB) in Taiwan put forward a professional health care quality index according to various overall quality schemes stipulated by the Department of Health for long-term monitoring. The main items for hemodialysis health care quality include: serum albumin, clearance rate of urea nitrogen (Kt/V), hematocrit, hospitalization rate, death rate, sinus reconstruction rate, and weaning rate. Then NHIB could evaluate the quality of dialysis centers' medical service based on these items.

Generally speaking, if a patient got uncomfortable feeling in the process of dialysis treatment, the nurse would enquire the patient and suggest the patient to take relevant examinations and treatments according to her professional practice, so as to avoid other infections. However, many diseases will not show obvious symptoms at usual times nor to be perceived easily. Whereas once these infections get worse, the patient needs a hospitalization for observation and treatments. Some serious case may result in death [12].

A relevant study to hemodialysis quality, Hong [13] used multi-mini-support association rules and frequent time serial algorithm for mining biochemical test data of hemodialysis patients. He found out patterns for predicting hospitalization. The multi-mini-support association rules therein are mainly for improving previous method of setting single support only, and setting different supports for different items. They work out coincident rules with actual situation. Finally this study found out some commonly known professional knowledge, as well as some patterns with clinical and preventive medicine.

B. Temporal Abstraction

Temporal abstraction (TA) is an artificial intelligence technique, which can integrate domain knowledge into the process of data analysis. It outlines the evolutionary process of temporal data through qualitative presentation mode, such as level shifts, periods of stability and trends. Shahar [2] defined the temporal abstraction as a program given a set of time series data including variables, external events, and abstract. The generated abstract description can represent previous and current states and trends of data. Based on TA program, patients' data can be converted from a low-level quantitative format to a high-level qualitative description. This presentation format is close to the vocabulary of clinician specialty [4]. Rules and knowledge based temporal abstraction are usually discussed by clinicians and domain experts together. These knowledge and rules are very important for generating significant and data-dependent temporal abstraction. They will determine whether these abstractions could be explained correctly to work out correct diagnoses. Generally, TA can be obtained from both basic TA and complex TA which are to be described respectively as follows.

1) Basic TA

Basic TA refers to qualitative description converted from time series data of existing episodes. It can be usually indicated by a combination of state and trend. The state can be classified as low, normal, and high values. The trend can be classified as increase, decrease, and stable patterns [2]. An episode refers to the state of data in a time interval. It is for providing a brief description of each variable's state in a certain time interval [5].

2) Complex TA

Complex TA describes the temporal relation between basic TAs or other complex TAs. In other words, basic TA is obtained before composing complex TA [5]. Typical complex TAs usually use temporal operator proposed by Allen [14] to concatenate basic TA (see Table I). The mostly used one is Meet, referring to successively presented precedence order between basic TAs [4]. For example, if a rule, two basic TAs A-high and B-high correspond result C, and A occurs earlier than B, it can be indicated by A-high Meet B-high Then C.

3) Application of temporal abstraction

Temporal abstraction is an indispensable element in intelligent data analysis (IDA). Initially it is applied to data monitoring of patients in intensive care units (ICU) based on the intelligent data analysis system which detects abnormal phenomena of patients' temporal data. Then the clinicians are provided with relevant temporal presentation information of patients for subsequent treatment actions. Lavrac et al. [15] defines intelligent IDA as "encompassing statistics, pattern recognition, machine learning, data abstraction and visualization tools to support the analysis of data and discovery of rules that are hidden in massive time series data." IDA system schema is as shown in Figure 1, containing several different

TABLE I
TEMPORAL OPERATOR [5]

Relation	Examples
A operator(PRECEDE) B	Overlap : aaaaa bbbbb
	Meet: aaaaa bbbbb
	Before : aaaaa bbbbb
	Equal : aaaaa bbbbb

missions: data validation, data representation, data interpretation, and control tasks. Temporal abstraction is applicable to data representation and interpretation phase. During the reasoning process, when TA is applied to the data analysis system, interpretation or reasoning can be carried out by comparing predefined patterns defined by clinicians or derived from machine learning techniques. Finally the control tasks present proper treatment actions according to abnormal phenomena of data.

The application of temporal abstraction is also related to medical data analysis, and often integrated with data mining techniques and artificial intelligence approaches. In the study of [5], the blood pressure and ferroheme of hemodialysis patients were derived from dialyzers in weekly three times of treatments. They were processed by discrete wavelet transformation and association rules mining to obtain the insufficient factors of hemodialysis service quality. Then the mined rules would be provided to clinicians for reference. The results showed that the context of an auditing system for dialysis management helped clinicians to improve their understanding of the patients' behavior. Ho et al. [16] integrated decision tree with temporal abstraction to analyze hepatitis patients' data. Differing from [5], the hepatitis data used in [16] was collected at irregular intervals. Therefore the time intervals of temporal abstraction were not the same. In addition, they used "state change" temporal operator to describe temporal abstraction, and used the formula counting significance to prune the rules. The experimental results show that the generating hepatitis temporal pattern was improved effectively.

C. Decision Tree

A decision tree (DT) is a flow-chart-like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node represents hospitalization or non-hospitalization. The top-most node in a tree is the root node. The test on an attribute is associated with a splitting criterion which is chosen to split the data sets into subsets that have better class separability, thus minimizing the misclassification error. Once the tree is built from the training data, it is then heuristically pruned to avoid over-fitting of data, which tends to introduce classification error on the test data [17].

This approach is a well-known tree-growing algorithm based on univariate splits called C4.5 that is an extended version of Quinlan's ID3. The original ID3 algorithm used a criterion called gain to select the gene to be tested which is based on the information theory concept, entropy, defined as:

$$Info(S) = entropy(P, 1-P) = -P\log_2(P) - (1-P)\log_2(1-P) \quad (1)$$

where S is a set of samples and P is the probability of a tuple being hospitalization. Assume that the task of selecting a possible test with n outcomes (numbers of some temporal abstractions of a certain testing item) that partitions the set T of training samples into subsets T1, T2,…, Tn. After T has been partitioned in accordance with n outcomes of one attribute text X. The expected information requirement can be found as the weighted sum of entropies over the subsets:

$$Info_x(T) = -\sum_{i=1}^{n}(|T_i|/|T|)Info(T_i). \quad (2)$$

The quantity Gain(X) measures the information that is gained by partitioning T in accordance with the test X. The gain criterion selects a test X to maximize Gain(X); i.e., this criterion will select an attribute with the highest information gain as follows:

$$Gain(X) = Info(T) - Info_x(T) \quad (3)$$

During the process of tree construction, this function is applied to find the best attribute to split and the best splitting criterion for the chosen attribute. This is done by testing each unused attribute on all of its possible splitting points, using the purity entropy function and selecting the one that gives the best result.

III. Development of Decision Support System

The system schema of this study is mainly developed from IDA, as shown in Figure 2. In the knowledge creation phase, firstly knowledge engineers and relevant professional persons of hemodialysis work together to delete outliers and handle missing data from time series data of dialysis patients. The knowledge of professional medical care personnel applied to temporal abstraction is obtained, and then the pre-processed data will be changed into temporal abstraction format based on TA rule obtained by the system. This process integrates the knowledge of specialized fields into analysis program. Afterwards, the transformed temporal abstraction data is fetched into the data mining program to find out rules for predicting hemodialysis patients' hospitalization. The mined patterns are then provided to clinicians for further processing including rules evaluation and irrational rules deletion. The remained rules are stored into database, so as to provide a network system as assistance for clinicians to practically judge whether patients need hospitalization or not.

Open Science Index, Economics and Management Engineering Vol:3, No:5, 2009 publications.waset.org/3884/pdf

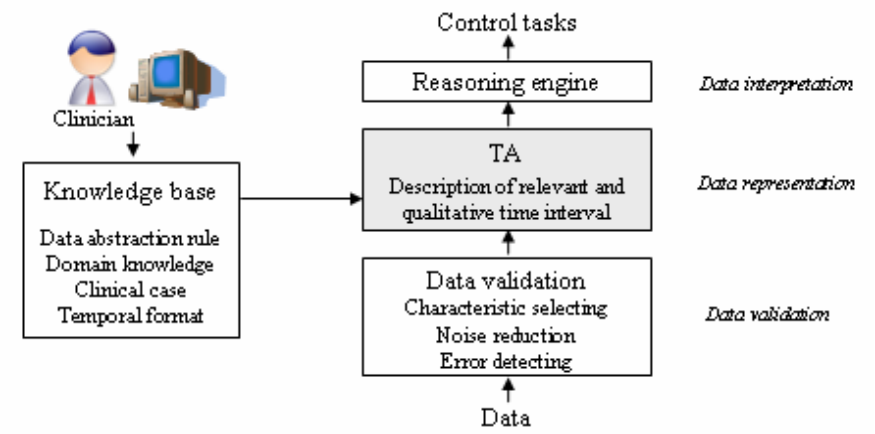


Fig. 1. Typical intelligent analysis system schema (Stacey & McGregor, 2007)

Open Science Index, Economics and Management Engineering Vol:3, No:5, 2009 publications.waset.org/3884/pdf

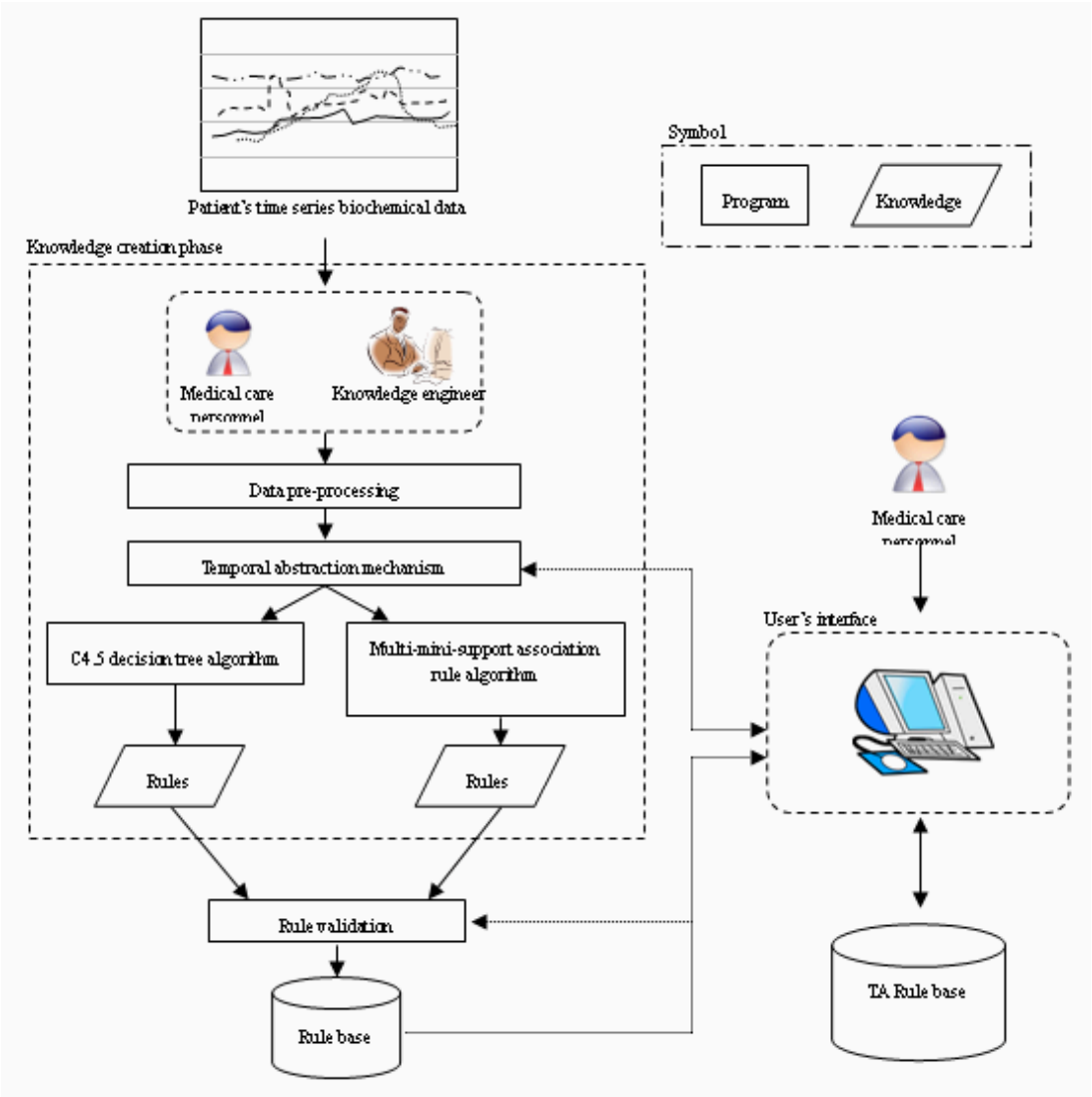


Fig. 2. Flow of hemodialysis patients' data analysis

The sample of this study is the time series data of patients obtained from a nationwide hemodialysis center, Taiwan, including monthly routine biochemical testing items, sex, age, and hospitalization records of patients.

Open Science Index, Economics and Management Engineering Vol:3, No:5, 2009 publications.waset.org/3884/pdf

A. Data Pre-processing

Due to heterogeneity and specialty problems of medical data, we analyzed the biochemical tests of original hemodialysis patients during the pre-processing phase. It is also necessary to cooperate with professional medical care personnel in this step to find out outliers and missing values.

1) Feature selection

From data mining point of view, every biochemical testing item represents a feature. Due to different biochemical testing items having their own clinical significance, we need to discuss with professional medical care personnel to sort out the applications of various biochemical testing items. Also the clinical significance of some biochemical testing items in data will appear frequently. After enquiry, we can select the feature and avoid redundant features influencing the accuracy and effectiveness of mining. In all biochemical testing items of hemodialysis patients' data in this study, for example, protein and albumin represent patients' nutritional status. The clinicians suggest selecting any of them as the feature.

2) Missing data processing

Considering difficulty to get medical data and the number of data volume, we discarded those features with missing data more than 15%, such as Ferritin, Tranferrin saturation, Albumin (Al), Cardiac/thoracic ratio, etc. For those features have missing data less than 15%, we filled up the mean value of the previous and latter values of the patient.

3) Outlier processing

Due to hemodialysis patients' physique is different from normal persons', a minority of patients' routine biochemical test values will be ultra-abnormal. This is often seen in medical data. Furthermore, when medical care personnel input biochemical test values into computers, it is hard to avoid typo. This will result in serious errors in data analysis. Therefore, it is also required to contact medical care personnel about the standard value scope of various testing items in this step. Then the outliers could be found based on three times of standard deviation [17].

4) Addition of derived variable

Various biochemical testing items of hemodialysis patients represent different clinical significance. The clinical significance will be clear because of new values of some items obtained after mathematical computation. For example, before-dialysis blood urea nitrogen (BUN) and after-dialysis BUN, if we put two test values into formula for computation [(before-dialysis BUN - after-dialysis BUN)/before-dialysis BUN], we will obtain urea reduction ratio (URR) value. In addition, the product of calcium and phosphorous (Ca×P) has a special significance on clinic. If serum inorganic phosphorus is greater than 5.5 and Ca×P is greater than 55, the death rate would increase accordingly (calcification of main coronary artery and cardiac muscle result in coronary heart disease and heart failure). Therefore the value of Ca×P needs additional calculations and to be added to the system.

B. Transforming to Temporal Abstraction

The steps of transforming to temporal abstraction are as shown in Figure 3. Corresponding normal standard interval value to each biochemical test value shall be obtained in advance, and then the transform of temporal abstraction could be carried out. Firstly transform patients' biochemical test value data into Basic TA through "basic TA algorithm", and then input the result into "complex TA algorithm" to figure out Complex TA.

The sample of this study is mainly the patients taking long-term hemodialysis treatments, according to the definition given by the National Health Insurance Bureau, Taiwan. A patient who has taken hemodialysis treatment for more than three consecutive months is a long-term dialysis patient. The National Health Insurance Bureau monitors the hospitalization rate once every six months. This study would take the patients who have taken hemodialysis treatment for more than six consecutive months as the analysis subject. Furthermore, Hong [13] used test data of three months to predict the possibility of patients' hospitalization. His study sets the time interval of basic TA at three months. However, in order to find out the evolutionary pattern of biochemical test value which may cause patients' hospitalization in six months of hemodialysis treatment, our study defines the complex TA as six months for one interval. If a long-term hemodialysis patient doesn't have any hospitalization within six to twelve months after treatment, it would be regarded as non-hospitalization data se. If the patient has hospitalization in six months after treatment, this would be regarded as hospitalization data set. We would get the complex TA of non-hospitalization data set patients in late six months, as well as the complex TA of hospitalization data set patients in six months prior to hospitalization.

1) Transforming to basic TA

The time series data of hemodialysis patients should be transformed to qualitative description mainly with time interval before data mining, so as to execute temporal mining. Firstly we need to define the format of basic TA according to the abstraction rules given by medical care personnel (as Figure 4):

For example, assuming it is the time series data of biochemical test values of a patient from January to October shown in Figure 5. According to the abstraction intervals of basic TA and complex TA, we can select a set of basic TA and a set of complex TA from the data.

2) Transforming to complex TA

Basic TA composing complex TA can be concatenated by temporal operator. The temporal operator, > (from one state to another), states transforming and trend evolution. This study

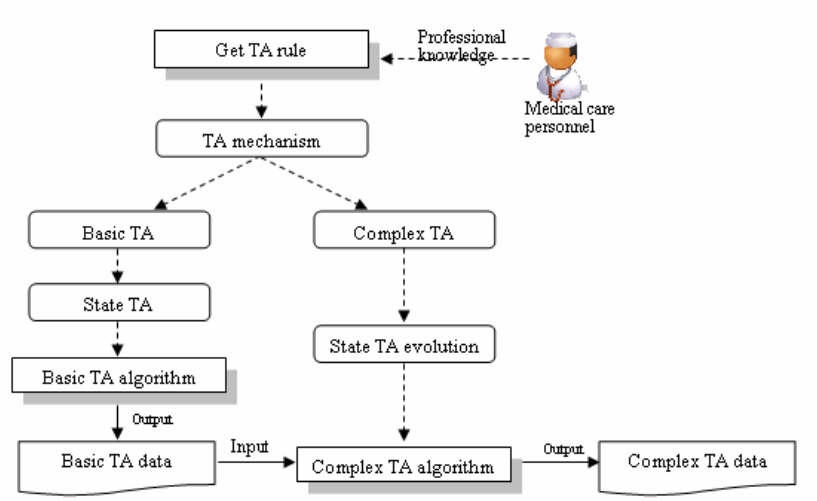


Fig. 3. Temporal abstraction flow (note: : mission; : concept; : data)

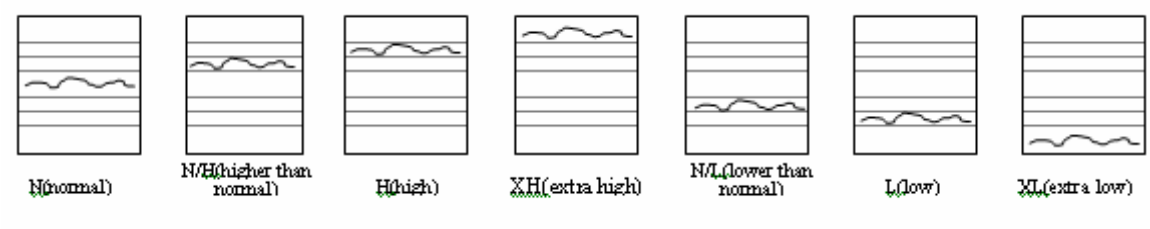


Fig. 4. Basic TA

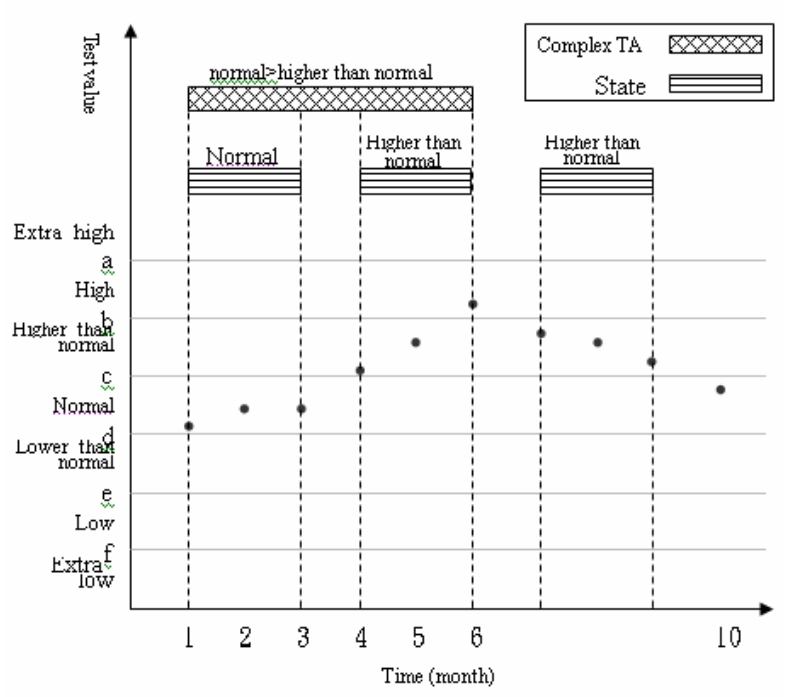


Fig. 5. Schematic diagram of temporal abstraction

uses the temporal operator put forward by [16] to concatenate basic TA. For example, the patient's long-term representation can be indicated by complex TA "normal>higher than normal (N>N/H)". We can see the patient's biochemical test value changes from normal to higher than normal state from the first month to the sixth month. It also means the patient's biochemical test value has a trend to climb. Therefore, the complex TA formed by using ">" temporal operator can

Open Science Index, Economics and Management Engineering Vol:3, No:5, 2009 publications.waset.org/3884/pdf

contain basic state and trend abstraction at the same time. After the basic TA is obtained, we can input all patients' basic TA data into the complex TA algorithm for transformation of temporal abstraction.

A. Hospitalization Rule Mining

This study is to integrate decision tree with multi-mini-support association rules to mine TA data of hemodialysis patients, and find out rules for predicting patients' hospitalization, so as to prevent hospitalization of hemodialysis patients.

1) Multi-mini-support association rule mining

There is a shortage when using Apriori for association rules mining. If the support is set too high, rules would be reduced. If the support is set too low, the quantity of rules would increase [18]. Therefore, Liu et al. [19] put forward stipulating different supports for items with different features based on the Apriori architecture. The result was actually more effective than Apriori that some important and infrequent item set association rules were found.

Dialysis patients have special physique, and some items have important index significance, such as clearance rate of urea nitrogen (Kt/V), URR, albumin and so on. These items' temporal abstractions are closely related to patients' survival rate. They are so important that the item's support must be particularly considered by professional medical care personnel. Based on the above description, different minimum supports shall be set for special testing items' temporal abstract through professional medical care personnel's expertise. Therefore, this study used multi-mini-support association rules mining algorithm to find out the most relevant temporal abstract of testing item causing hospitalization. MIS (minimum item support) represents the minimum support for each testing item defined as follows:

$$MIS(i) = \frac{TA(i) \cup H}{N} \tag{4}$$

where i represents testing item, TA(i) represents the temporal abstraction of each testing item, H is the number of hospitalization, and N is the number of patients. If a TA's MIS is less than threshold 0.1, then the TA's MIS was set at 0.1. To evaluate the correlativity between testing item and hospitalization, Conf is defined as follows [5].

$$Conf = \frac{TA(i) \cup H}{T} \tag{5}$$

where T is the frequent item set of testing items. This study used the evaluation rule statistical significance method proposed by [16] to delete rules. Assuming a rule is R:A → H, and confidence is Conf. Therein, p(H) is the probability of patient's hospitalization, nA is the number of items for hospitalization. If Conf = p(H), then this rule would be deleted [20]. Z value is namely the threshold of delete rule defined as:

$$Z = \frac{Conf - p(H)}{\sqrt{p(H)(1 - p(H))/n_A}} \tag{6}$$

IV. Experimental Results and Analysis

The experimental results of all steps and an evaluation method used in this study will be introduced successively in this section.

A. Data Format

Samples used in this study are 2005-2007 quarterly report data of patients from in a nationwide hemodialysis center. After sorting out these data, there are 8223 samples in all. After deleting those containing nulls and missing data as well as outliers, there are 6284 remaining samples.

B. Hemodialysis Knowledge System

In order to change the monthly biochemical test data of hemodialysis patients into temporal abstractions, we worked out the conversion rules for temporal abstraction of various biochemical testing items. This study develops a system to extract professional persons' hemodialysis knowledge, and presents the results of data mining in the research to domain experts for evaluation. Furthermore, it provides assistance to clinicians for practical applications.

1) Conversion rules for temporal abstraction

There is no efficient way for clinicians to share about professional knowledge face to face due to their busy works, such as recording patients' biochemical values and replacement and setting of hemodialysis devices and so on. Therefore, this system implements a master-slave architecture based on network to obtain professional knowledge by visualization. The threshold of conversion rules for temporal abstraction inputted by clinicians is presented by visual mode. They can also indicate the clinical significance of the biochemical testing item in the remarks column.

2) Hospitalization pattern mining

This study mainly composed of C4.5 decision tree and MSApriori multi-mini-support association rule algorithm. The complex TA obtained will be inputted into these two algorithms for hospitalization pattern mining.

The result of C4.5 decision tree determines which items can be used as the index for predicting hospitalization mainly by calculating the information gain of each biochemical testing item. Before data mining, this study uses the feature selecting function of Weka software to check out the significance level of each biochemical item [21]. We use InfoGainAttributeEval as the index for calculating significance level of each biochemical item, and use Ranker method to sequence all biochemical items according to their significance levels. There are 25 biochemical items in all. When biochemical items to be analyzed are selected, we input them into C4.5 decision tree for construction. Most of parameter settings are presetting of software, and confidence factor for trimming is changed from 0.25 to 0.3.

Open Science Index, Economics and Management Engineering Vol:3, No:5, 2009 publications.waset.org/3884/pdf

Open Science Index, Economics and Management Engineering Vol:3, No:5, 2009 publications.waset.org/3884/pdf

MSApriori requires patients' complex TA data set, output data set, support data set of each item, and support (Conf = 0.7) for association rule mining. This software is not applicable to processing data of item scale. Therefore, all biochemical items should be transformed to code data, so as to proceed with association rule mining.

C. Result Evaluation

As for the evaluation of experimental results, this study puts forward different evaluation methods for decision tree and association rule respectively.

1) Result evaluation of decision tree

This study uses 10-fold cross validation to validate C4.5 decision tree model. In addition, Podgorelec et al. [11] indicated that in medical data analysis region, it is appropriate for accuracy of classification predicting to use sensibility and specificity to evaluate prediction model. The formulae are as shown in Table II. Sensibility is mainly for evaluating whether the decision tree could predict the hospitalization rate of patients correctly; specificity is for evaluating whether the decision tree could predict non-hospitalization rate of patients correctly. The main purpose of this study is to find out biochemical test value patterns causing patients' hospitalization. Therefore, the evaluation of sensibility will be emphasized.

Figure 6 is a comparison of time intervals of different temporal abstractions to see whether there will be difference in results or not. For hospitalized patients, in normal conditions, if a patient has hospitalization record in July, then the biochemical value record from January to June would be obtained. This is the test group. Then we take the same patient's biochemical value record from February to July as the control group. From the figure, we can know the sensibility of time interval of test group is much higher than that of control group. The correctness and rationality of this study on transforming temporal abstraction are proved.

Figure 7 is the comparison between impacts of different abstraction rules on research results. Minor-abstraction rule means the biochemical test values are divided into three temporal abstractions (high, normal, low); multi-abstraction rule means the biochemical test values are divided into seven temporal abstractions (extra high, high, higher than normal, normal, lower than normal, low, extra low). From the result we can see the overall accuracy rate of three temporal abstractions is higher, but this study lays emphasis on accuracy of sensibility. Therefore, the result of seven temporal abstractions is better than that of three temporal abstractions.

2) Result evaluation of association rule

The support of hospitalization association rule mining in this study will set 0.2 initially and adjust the support with 0.05 interval value. The unnecessary rules are deleted based on the rule reducing formula of [11]. Finally, after the rules are verified by domain experts, we determine a support value with more clinical significance.

TABLE II
PERFORMANCE EVALUATION INDEX FORMULAE

	(classified as) hospitalization	(classified as) non-hospitalization
(actual) hospitalization	TP	FN
(actual) non-hospitalization	FP	TN
sensitivity	$\frac{TP}{TP+FN}$	
	TN	

D. Results of Data Mining

After the analytic results of this study were evaluated by domain experts, rules were obtained to predict hemodialysis patients' hospitalization effectively. They are listed as follows for interpretation.

1) Results of decision tree

Results of using decision tree mining are as follows, group presentation and interpretation are made based on testing items. The end of each rule in brackets is the value of accuracy rate for predicting patients' hospitalization through that rule.

a) Relevant rule to albumin

IF (Albumin = "normal(N) > extra low(XL)") THEN (Hospitalization = YES) (93%)

Albumin always represents current nutrition status and inflammation degree. It is one of upmost factors influencing long-term survival. If the albumin is less than 3.4 gm/dl, the patient was likely to be infected due to malnutrition, and the hazard of death would increase accordingly. If the albumin is less than 2.5 gm/dl, the death relative risk would increase by 16 times, and relative hospitalization rate would also be high [1].

b) Relevant rule to albumin, hemachrome (Hbc), and platelet

IF (Albumin = "lower than normal(N/L) > lower than normal(N/L)") AND (Hbc = "normal(N) > low(L)") THEN (Hospitalization = YES) (100%)

Patient with low Hbc or insufficient Hbc indicates that he has anemia. More worse when he has malnutrition, he may have insufficiency of antibody and is likely to get other diseases. The major function of platelet is forming thrombosis and concretionary shrinkage when arresting bleeding. The value of platelet increases because of anemia, hepatocirrhosis, acute infection, chronic granular leukaemia, rheumatiod arthritis and so on.

c) Relevant rule to albumin, triglyceride, and age

IF (Albumin = "lower than normal(N/L) > low(L)") AND (Triglyceride = "lower than normal(N/L) > low(L)") AND (age > 74) THEN (Hospitalization = YES) (100%)

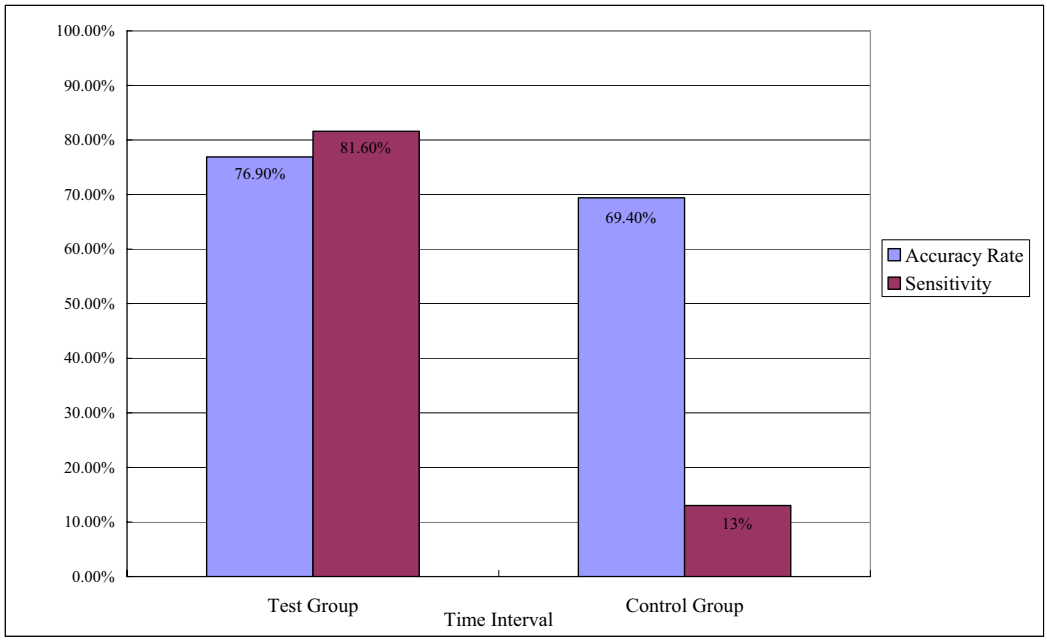


Fig. 6. Comparison between results of different time intervals sampling

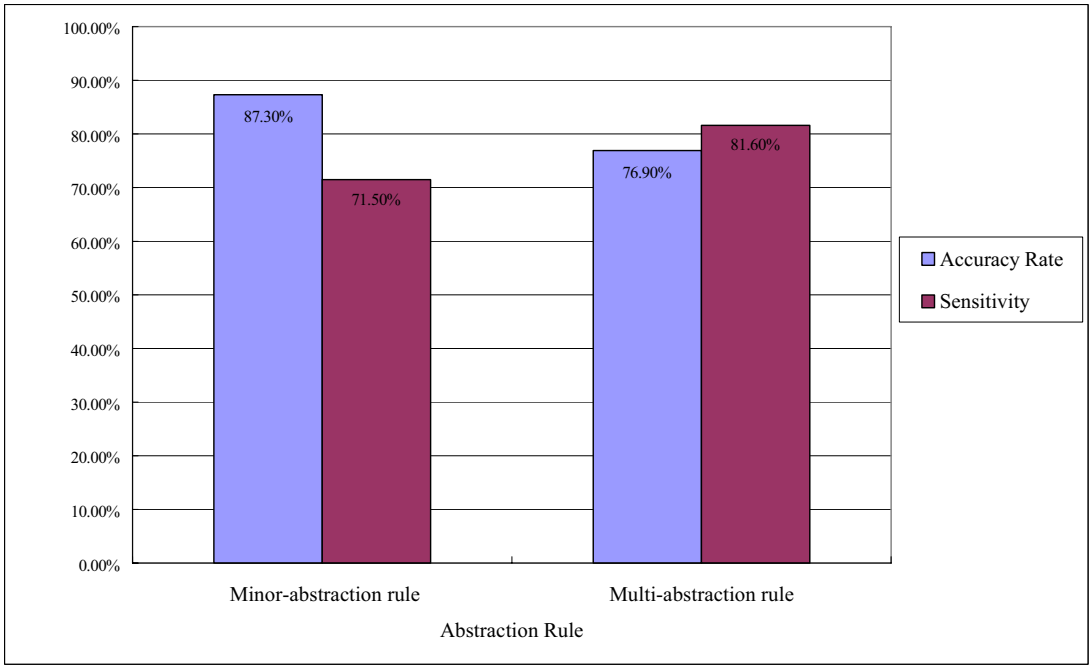


Fig. 7. Comparison between results of different temporal abstraction rules

Triglyceride is the fat in blood and mainly provides cellular capacity. Decline of triglyceride may cause lack of lipoprotein, chronic obstructive pulmonary disease, hyperthyroidism and malnutrition. The aged hemodialysis patients are likely to get hospitalization because of insufficient ingestion of nutrition and low triglyceride.

2) Results of association rule mining

a) Rules related to blood examination

IF (Hct = "normal(N) > lower than normal(N/L)") THEN (Hospitalization = YES) (Minsup:0.079; Conf:0.77)

Decrease of Hct is an index for judging whether the patient is anemic or not. Patient with lower Hct will cause anoxia, dyspnea, and dilutedness which are symptoms of serious

Open Science Index, Economics and Management Engineering Vol:3, No:5, 2009 publications.waset.org/3884/pdf

Open Science Index, Economics and Management Engineering Vol:3, No:5, 2009 publications.waset.org/3884/pdf

anemia.

IF (Erythrocyte = "lower than normal(N/L) > low(L)") AND (Albumin = "normal(N) > lower than normal(N/L)") THEN (Hospitalization = YES) (Minsup:0.013; Conf:0.87)

Too low erythrocyte value means anemia. The patient shall use erythropoietin timely. Otherwise with insufficient ingestion of nutrition, he is likely to get other diseases.

b) Other rules related to examination value

IF (Diabetes = high) AND (Ca×P = "lower than normal(N/L) > lower than normal(N/L)") AND (Phosphorous = "normal(N) > normal(N)") THEN (Hospitalization = YES) (Minsup:0.02; Conf:0.93)

Generally speaking, the higher Ca×P is, the more likely coronary heart disease and heart failure to be caused by calcification of main coronary artery and cardiac muscle. However, maybe diabetes patients have special physique, the standard interval of Ca×P shall be discussed additionally for diabetes patients.

If (Albumin = "normal(N) > lower than normal(N/L)") THEN (Hospitalization = YES) (Minsup:0.065; Conf:0.83)

The same result as decision tree, over low albumin is still an important index for prognosis.

c) Presentation of hospitalization rule

All significant rules validated by domain experts will be stored in the rule base of this system. The mined rules are presented through network inquiry system, for clinicians to inquire, so as to assist them to judge the possibility of hospitalization.

V. Conclusion

The process of medical data analysis includes data gathering, pre-processing, result evaluation, favorable interaction, and discussions with professional health care personnel to make the analytic results correct. This study uses network database system for extracting professional knowledge. This improves the traditional face-to-face talking with professional persons, and enables us to obtain important knowledge effectively and quickly.

From the experimental results, we can see that data mining methods for different purposes can be combined effectively, and more abundant patterns can be found out for practical applications. Furthermore, we can add domain knowledge prior to data analysis by combining temporal abstraction method, to make mining results more likely to be comprehended by professional medical care personnel. Therefore, for future medical time series data analyses, the temporal abstraction will be an indispensable method.

Among the hospitalization patterns found out by this study, the albumin is the most important index for predicting patients' hospitalization. This index is also used clinically for predicting patients' death rate at present. Therefore, it is proved that the results of this study have clinical significance. In addition, predicting patients' hospitalization by biochemical value evolution of blood examination was an undefined biochemical item in previous clinical applications. After validated by medical care personnel, the time evolution of this index value is proved to have definite relevance with hospitalization.

In this study, we combined temporal abstraction method with data mining techniques to analyze the dialysis patients' biochemical data. The mined temporal hospitalization patterns will be helpful for doctors to diagnose hospitalization probability of patients and to suggest them some treatments immediately to avoid hospitalization. Finally, we hope we could help the hemodialysis centers to improve their health care quality through this research.

There are still many relevant methods and concepts can be added in continuously for analysis of the results. We list two points of researchful directions in later day for scholars' reference. Adlassnig et al. [3] indicates that combining temporal abstraction method with Fuzzy Logic theory is more coincident with actual situation for description of temporal data. It is also a direction worth study.

Acknowledgment

We are grateful to the National Science Council for the grants program (NSC 96-2221-E-415-005), and to the Hemodialysis Center of Chiayi Yang Ming Hospital for their professional consultations.

References

[1] C. G. Tan. (2004) Interpretation of periodic examination results of dialysis patients. [Online]. Available: http://www.kidney.org.tw/doc/49/01.doc

[2] Y. Shahar, "A framework for knowledge-based temporal abstraction," *Artificial Intelligence*, vol. 90, pp. 79-133, 1997.

[3] K. P. Adlassnig, C. Combi, A. K. Das, E. T. Keravnou, and G. Pozzi, "Temporal representation and reasoning in medicine: research directions and challenges," *Artificial Intelligence in Medicine,* vol. 38, pp. 101-113, 2006.

[4] M. Stacey, and C. McGregor, "Temporal abstraction in intelligent cinical data analysis: A survey," *Artificial Intelligence in Medicine*, vol. 39, pp. 1-24, 2007.

[5] R. Bellazzi, C. Larizza, P. Magni, and R. Bellazzi, "Temporal data mining for the quality assessment of hemodialysis services," *Artificial Intelligence in Medicine*, vol. 34, pp. 25-39, 2005.

[6] U. Fayyad, G. P. Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI Magazine,* vol. 17, no. 3, pp. 37-54, 1996.

[7] R. Bellazzi, and B. Zupan, "Predictive data mining in clinical medicine: current issues and guidelines," *International Journal of Medical Informatics*, vol. 77, no.2, pp. 81-97, 2008.

[8] A. Kusiak, B. Dixon, and S. Shah, "Predicting survival time for kidney dialysis patients: A data mining approach," *Computers in Biology and Medicine*, vol. 35, pp. 311-327, 2005.

[9] J. Demsar, B. Zupan, N. Aoki, M. J. Wall, T .H. Granchi, and J. R. Beck, "Feature mining and predictive model construction from severe trauma patient's data," *International Journal of Medical Informatics,* vol. 63, pp. 41-50, 2001.

[10] M. J. Huang, M. Y. Chen, and S. C. Lee, "Integrating data mining with case-based reasoning for chronic," *Expert Systems with Applications*, vol. 32, pp. 856-867, 2007.

[11] V. Podgorelec, P. Kokol, M. M. Stiglic, M. Hericko, and I. Rozman, "Knowledge discovery with classification rules in a cardiovascular dataset," *Computer Methods and Programs in Biomedicine,* vol. 80, pp. S39-S49, 2005.
[12] H. H. Feng. (2005). Prevention and treatment of dialysis patients' heart failure. [Online]. Available: http://www.kidney.org.tw/doc/51/03.doc
[13] G. Q. Hong, "Multiple-minimum-support association rule mining for hospitalization prediction of hemodialysis patients," M.S. thesis, Department of Information Engineering, National Tung Hua University, Taiwan, 2004.
[14] J. F. Allen, "Towards a general theory of action and time," *Artificial Intelligence*, vol. 23, pp. 123-154, 1984.
[15] N. Lavrac, E. Keravnou, and B. Zupan, "Intelligent data analysis in medicine," In: Kent A., editor, *Encyclopedia of Computer Science and Technology*, vol. 42, New York, USA, Basel: Marcel Dekker, 2000, pp. 113-57.
[16] T. B. Ho, T. D. Nguyen, S. Kawasaki, S. Q. Le, D. D. Nguyen, H. Yokoi, and K. Takabayashi, "Mining hepatitis data with temporal abstraction," *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Washington, DC, USA, 2003.
[17] M. Kantardzic, *Data Mining: Concepts, Models, Methods, and Algorithms*, John Wiley, New York, 2003.
[18] Y. H. Ding, M. Y. Chen, *Data Mining*, Canghai Books, Taiwan, 2003.
[19] B. Liu, W. Hsu, and Y. Ma, "Mining association rules with multiple minimum supports," *ACM SIGKDD International Conference on knowledge Discovery & Data Mining (KDD-99)*, San Diego, CA, USA, 1999.
[20] T. Takabayashi, T. B. Ho, H. Yokoi, T. D. Nguyen, S. Kawasaki, and S. Q. Le, "Temporal abstraction and data mining with visualization of laboratory data," *Medinfo 2007 Congress*, Brisbane, pp. 1304-1308, 2007.
[21] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd Ed., Morgan Kaufmann, New York, USA, 2005.

Jinn-Yi Yeh received the B.S. degree from Chung-Yuan Christian University, Taiwan, in 1987 and the M.S. and Ph.D. degrees from the University of Texas at Arlington in 1993 and 1997, respectively. He is presently an Associate Professor in the Department of Management Information Systems, National Chiayi University, Taiwan. His research interests focus on data mining, medical image processing, and heuristic algorithms.

Tai-His Wu is a professor in the Department of Business Administration, National Taipei University. He received his master and Ph.D. degrees in Industrial Engineering from Texas A&M University, USA, in 1994. His research and teaching interests include developing models and algorithms for production and transportation problems.

Open Science Index, Economics and Management Engineering Vol:3, No:5, 2009 publications.waset.org/3884/pdf