MULTIPLE CRITERIA DECISION MAKING
Vol. 10 2015

Chao-Wen Chen[*]
Yuh-Wen Chen[**]
Moussa Larbani[***]
Tzung-Hung Li[****]

USING MULTI-OBJECTIVE AFFINITY MODEL FOR MINING THE RULES OF REVISITS WITHIN 72 HOURS FOR EMERGENCY DEPARTMENT PATIENTS

Abstract

When patients return to the emergency department (ED) within 72 hours after their previous ED discharge, it is generally assumed that their initial evaluation or treatment had been somehow inadequate. Mining data related to unplanned ED revisits is one method to determine whether this problem can be overcome, and to generate useful guidelines in this regard. In this study, we use the receiver operating characteristic (ROC) curve to compare the data mining model by affinity set to other well known approaches. Some scholars have validated the affinity model for its simplicity and power in handling information systems especially when showing binary consequences. In experimental results, SVM showed the best performance, with the affinity model following only slightly behind. This study demonstrated that when patients visit the ED with normotensive status or smooth breath patterns, or when the physician-patient ratio is moderate, the frequency with which patients revisit the ED is significantly higher.

Keywords: Revisit, Emergency Department (ED), Data Mining, Affinity Set, Multi-objective.

[*] Kaohsiung Medical University Hospital, Division of Trauma, Department of Surgery, and Department of Emergency Medicine, 100 Tzyou 1st Rd., San Ming District, Kaohsiung City, Taiwan.
[**] Da-Yeh University, Institute of Industrial Engineering and Management, 168 University Rd., Da-Tsuen, Chang-Hwa 51505, Taiwan, e-mail: profchen@mail.dyu.edu.tw, tel. +886(4)8511888 ext 4120, fax +886(4)8511270.
[***] IIUM University, Department of Business Administration, Kulliyyah of Economics and Management Sciences, Jalan Gombak, 53100 Kuala Lumpur, Malaysia.
[****] Da-Yeh University, Institute of Industrial Engineering and Management, 168 University Rd., Da-Tsuen, Chang-Hwa 51505, Taiwan.

1 Introduction

Emergency physicians are expected to diagnose diseases accurately and efficiently. However, in fast-paced situations, time limitations and dynamic changes in the number of patients awaiting treatment lead to the inevitable risk of diagnostic error, by the simple fact that seemingly insignificant symptoms can be overlooked (Aaland, Smith, 1996; Brooksa, Holroydb, Riley, 2004; Kohn, Corrigan, Donaldson, 1999; Leape et al., 1991). Ignorance of such details could lead to a higher frequency of patients revisiting emergency rooms. Because emergency departments (EDs) are required to assume ever greater responsibilities, public interest in the quality of service they provide is increasing (Furnival, Woodward, Schunk, 1996; Hanlon, Pickette, 1979). Unscheduled revisits to EDs are known as audits of emergency care quality. Unscheduled revisits are commonly defined as patients presenting for the same chief complaint within 72 hours of discharge from the ED. A rate of less than 1% has been proposed as acceptable quality care (Wu et al., 2008). Unscheduled revisits are a reflection of ED performance, and the underlying causes must be investigated. A number of doctors have proposed traditional statistical methods to deal with this issue. Pierce et al. (1990) began an investigation into this important issue in 1990, followed by Hu (1992), Gordon et al. (1998). Recently, Wu et al. (2008) used the categorical analysis of patient revisits to the emergency department, in which age, sex, final discharge, reason for revisit, and the symptoms of most common complaints were calculated from 34714 records. Nuñez et al. (2006), studied 250 cases and 250 controls from the ED. The measured outcomes were unscheduled returns, post-ED destination, and patient dissatisfaction. They concluded that unscheduled returns were associated with medical errors in prognosis, treatment, follow-up care, and information. Marcantonio et al. (1999) performed a matched case control study among patients who had been admitted to an academic hospital in a Medicare managed care plan. The patients were aged 65 years or older and had been readmitted to the hospital as emergency cases within 30 days of discharge. They suggested that interventions, such as improved discharge education programs, could reduce unplanned readmission. However, most of the above studies applied traditional categorical analysis to the statistics, and tended to agree that revisits are generally illness-related. Further studies are needed to identify the most common and the most serious contributing factors related to revisits, to determine whether improvements can be made.

Early in 2004, Freitas (2004) reviewed the basics of multi-objective optimization for data mining, and suggested these optimization techniques are appropriate in data mining. Recently in 2012, Corne et al. (2012) proposed similar ideas for integrating multi-objective programing in supporting vector machines

(SVMs) (Cortes, Vapnik, 1995), decision trees (Abu-Hanna, Keizer, 2003), neural networks (Zbikowski, Hunt, ed., 1996) etc. These previous efforts validate the feasibility of using multi-objective optimization for mining big data. However, there are still limited multi-objective applications devoted to this area in addition to the popular evolutionary/soft methods (Freitas, 2008).

In this study we eschewed traditional statistical analysis, and employed a number of popular data mining techniques (Aguilar-Ruiz, Costa, Divina, 2004; Berman, 2002; Grupe, Owrang, 1995) to analyze collected clinical data of EDs rather than evolutionary/soft approaches. We adopted neural networks (Zbikowski, Hunt, 1996), rough sets (Rosetta) (Pawlak, 1991), SVM, decision trees, association rules (Delgado et al., 2001) and logistic regressions (Collett, 2003; Delen, Walker, Kadam, 2005). All of them are applied to uncover the relationship between causes and consequences of ED revisits. The affinity models has been validated/tested by a number of scholars (Alanazi, Abdullah, Larbani, 2013; Chen et al., 2009; Esfandiaria et al., 2014; Larbani, Chen, 2009; Michnik, Michnik, Pietuch, 2008; Paoin, 2011; Wu et al., 2009) in the areas of medicine and finance. In this study, a multi-objective affinity model was originally proposed to construct the *k*-core, presenting a number of advantages over the other data mining models evaluated in this study.

This paper is organized as follows: Section 2 introduces the basic concepts and definitions of affinity sets, and proposes the basic data-mining model of affinity. Section 3 reviews the popular data mining models and summarizes their advantages and disadvantages. Section 4 presents the multi-objective affinity model of data mining. Section 5 takes the actual samples of revisiting patients from Kaohsiung Medical University Hospital of Taiwan, to validate the data mining concept using our multi-objective affinity model, to identify the key factors in the high frequency of patient revisits. In addition, we compare the performance of multi-objective affinity model and other popular data mining models, according to the receiver operating characteristic (ROC) curve (Zweig, Campbell, 1993). Finally, in Section 6, we present our conclusions and recommendations based on the data mining results.

2 Preparation for Study

First, we review the basic concepts and definitions of affinity, as well as its potential use in data mining (Chen et al. 2009; Larbani, Chen, 2009; Michnik, Michnik, Pietuch, 2008; Wu et al., 2009). Interestingly, the word of affinity is popularly used in the chemical/medical/social field with various definitions. In chemical physics, chemical affinity is the electronic property by which dissimilar chemical species are capable of forming chemical compounds (Matejtshuk,

1997). In medicine, affinity is mentioned with various biomedical definitions, such as affinity membranes for the removal of endotoxins (Wei et al., 2002) and the immune system (Achenbach et al., 2004). A number of scholars have applied the biometric concept to soft computing where they used the affinity function to develop artificial immune systems (Hunt, Cooke, 1995). In social sciences, scholars give affinity a different meaning: affinity is characterized by high levels of intimacy and sharing, usually in similar groups, also known as affinity groups (Cattell, 2001; Ve-McConnell, 1999). Marketing managers believe that people are more likely to buy brands that affinity groups like. In this manner, they are able to track consumer behaviour according to the social interaction of affinity (Zinkhan, 2002).

Based on the various definitions of affinity given above, we concluded that no formal framework or theory dealing with affinity as a unified concept have been developed, and few researchers have discovered that the basic idea of affinity could be used to provide models valuable in information sciences. Fuzzy set theory is among the best tools for representing vague and imprecise concepts (Zadeh, 1965); however, a type of membership function is necessary in fuzzy sets. In this paper, we use the well known concept of closeness or distance between any two objects in topology to represent affinity and develop a data mining model. Due to its general nature, this new relationship theory, affinity set theory, is able to describe the degree of similarity between objects, and represent general relationships between objects, such as closeness, belongingness, equivalence, which enable decision makers to use this simple concept for modeling. The affinity set theory has been recently introduced in (Larbani, Chen, 2009). For further details we refer the reader to (Larbani, Chen, 2009).

2.1 Basic Definitions

We introduce the definition of an affinity set.

Definition 2.1
An affinity set consists of any two object (real or abstract) that create affinity.

Definition 2.2
Let e be a subject and A an affinity set. Let W be a subset of X⊆U. The affinity between e and A is represented by the function:

$$aff_{\mathrm{A}}^{e}(\,.\,): \mathrm{W} \rightarrow [0,1] \qquad (1)$$
$$w \rightarrow aff_{\mathrm{A}}^{e}(w)$$

The value $aff_{\mathrm{A}}^{e}(w)$ expresses the degree of affinity between subject e and affinity set A with respect to variable w. When $aff_{\mathrm{A}}^{e}(w) = 1$ this means that the affin-

ity degree of *e* with affinity set A is at the maximal level with respect to variable *w*; but $aff_A^e(w) = 1$ does not mean that *e* belongs to A, unless the affinity measure $aff_A^e(w)$ is the degree of belongingness. When $aff_A^e(w) = 0$ this means that *e* has no affinity with A with respect to variable *w*. When $0 < aff_A^e(w) < 1$, this means that *e* has partial affinity with A with respect to *w*. Here we emphasize the fact that the notion of affinity is more general than the notion of membership or belongingness: the latter is just a particular case of the former.

Definition 2.3
The universal set, denoted by U, is the affinity set representing the fundamental principle of existence. We have:

$$aff_U^e(\,.\,): U \rightarrow [0,1] \qquad (2)$$
$$w \rightarrow aff_A^e(w)$$

and $aff_U^e(w) = 1$, for all existing objects with respect to *w*.

In other words the affinity set defined by the affinity "existence" has complete affinity with all previously existing objects, that exist in the present, and that will exist in the future. In general, in real-world situations, a traditional referential set S, such that for objects *e* not in *S*, $aff_A^e(w) = 0$ for all $w \in W$, can be determined. In order to make the notion of affinity set operational and for practical reasons, in the remainder of the paper, instead of dealing with the universal set U, we only discuss affinity sets defined on a traditional referential set S. Thus, in the remainder of the paper when we refer to an affinity set, we assume that sets S and W are given.

Definition 2.4
Let A be an affinity set. Then the function defining A is:

$$F_A(.,.): S \times W \rightarrow [0,1] \qquad (3)$$
$$(e, w) \rightarrow F_A(e, w) = aff_A^e(w)$$

An element in real-life situations often belongs to a set for some variables and does not for other variables. Such behavior can be represented using the notion of an affinity set. The behavior of affinity set A over time can also be investigated through its function $F_A(.,.)$.

Interpretation 2.1

i) For a fixed element *e* in S, the function (3) which defines affinity set A reduces to the fuzzy set describing the variation of the degree of affinity of the element *e* over variable *w*.

ii) For a fixed w, the function (3) reduces to a fuzzy set defined on S that describes the affinity between elements S and affinity set A with respect to variable w. Roughly speaking, it describes the shape or "content" of affinity set A with respect to w.

iii) In addition to **i)** and **ii)**, we cannot say or check that an affinity set is a special fuzzy set, unless we can prove that any affinity set A is contained in a fuzzy set B, and vice versa.

Definition 2.5

Let A be an affinity set and $k \in [0,1]$. We say that an element e is in the t-k-Core of affinity set A with respect to w, denoted by w-k-Core(A), if $aff_{\mathrm{A}}^{e}(w) \geq k$, that is:

$$w-k-Core(\mathrm{A}) = \{e \,|\, aff_{\mathrm{A}}^{e}(w) \geq k\} \tag{4}$$

when $k = 1$, w-k-Core(A) is called simply the core of A with respect to w, denoted by w-Core(A). In addition, w-k-Core(A) $\equiv$ k-Core(A(w)).

Definition 2.6

A life range is defined as the continuous or discrete mapping from the behavior of an element e of S to an affinity set A with respect to w: an illustration of the continuous case is given in Figure 1 below. However, a discrete case for v is also possible.

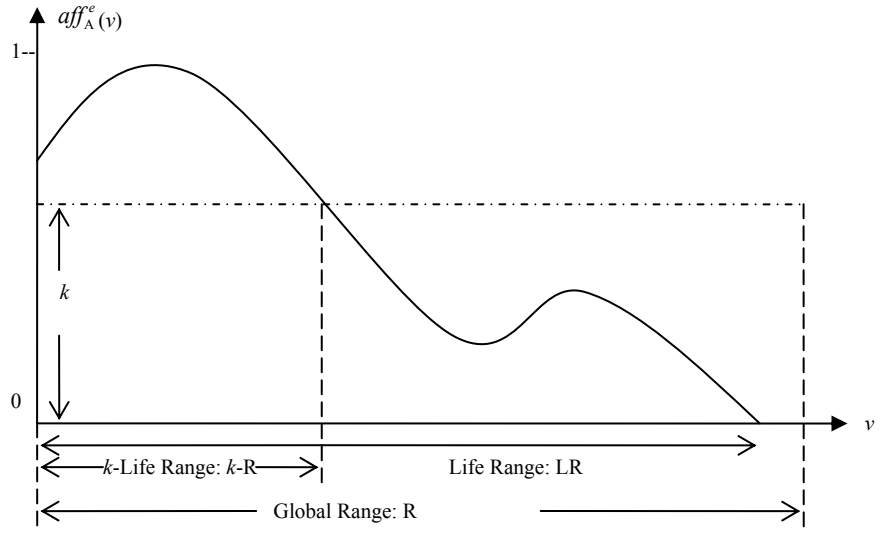

Figure 1. Illustration of the affinity between an element e and an affinity set A over a Global Range R (Continuous Case of v)

Here *k*-life range is the variable set: $\{v|$ for all $v \subseteq w$ such that $aff_A^e(v) \geq k\}$; similarly, life range is the variable set $\{v|$ for all $v \subseteq w$ such that $aff_A^e(v) \geq 0\}$.

The intersection and union operations on affinity sets are defined as follows.

Definition 2.7

The intersection of affinity sets A and B with respect to variable *w*, denoted by $A \cap B$, is defined by the function $F_{A\cap B}(e, w) = aff_{A\cap B}^e(w) = \text{Min}\{aff_A^e(w), aff_B^e(w)\}$, for all *e* in S. If A and B are considered over W, then $A \cap B$ is defined by the function:

$F_{A\cap B}(e, w) = aff_{A\cap B}^e(w) = \text{Min}\{aff_A^e(w), aff_B^e(w)\}$, for all *e* in S and all $w \in W$.

Definition 2.8

The union of A and B with respect to variable *w*, denoted by $A \cup B$, is defined by the function $F_{A\cup B}(e, w) = aff_{A\cup B}^e(w) = \text{Max}\{aff_A^e(w), aff_B^e(w)\}$, for all *e* in S. If A and B are considered over W, then $A \cup B$ is defined by the function $F_{A\cup B}(e, w) =$ $= \text{Max}\{aff_A^e(w), aff_B^e(w)\}$, for all *e* in S and all $w \in W$.

2.2 Affinity Data Mining

A static data mining model is proposed by using the basic theory of affinity.

Definition 2.9. Let V be a referential set endowed with distance $d(x, y)$, i.e. (V, d) is a metric space (Chen, 2009). Let X be a subset of V. The affinity set A in X is given by:

$$A = (d', B, X)$$

where d' is defined by:

$$d' : X \to [0,1]$$

$$e \to d'(e, B) = 1 - \alpha d(e, B)$$

where d' is the affinity, the set B is called the core of the affinity set A, $d(e, B)$ is defined by:

$$d(e,B) = \min_{z \in B} d(e, z)$$

Note that there is a difference between $d(e, B)$ and $d(x, y)$, although the same notation "*d*" is used. Indeed, $d(e, B)$ is the distance between an element *e* of X and the subset B of X, while $d(x, y)$ is the distance between two elements *x* and *y* of X. Note, that these two notions are different. Let $\alpha = \frac{1}{\max_{(x,y)\in X\times X} d(x, y)}$, that is, α is the inverse of the maximal distance between elements of X.

Procedure 2.1

1) Define the affinity set A, determine the referential set V and define the metric space (V, d).
2) Determine the set X.
3) Choose a subset B of X which is a candidate for being the core of the affinity set A.
4) Use the affinity d' defined by:

$$d' : \mathrm{V} \to [0,1]$$

$$e \to d'(e, \mathrm{B}) = 1 - \alpha d(e, \mathrm{B})$$

to compute the k-core (A) when, once the value of k is given. Now we present an example illustrating how this idea works.

Example 2.1. Data Mining

Table 1: Sample Data of Patients

Sample	x_1 (Fever)	x_2 (Vomiting)	y (Death)
P_1	0	1	1
P_2	1	0	1
P_3	1	0	0
P_4	0	1	1
P_5	1	0	0

Here we assume that doctors have observed two symptoms for one new disease: one is "Fever", the other is "Vomiting", and they possibly lead to the death of patients. We collect the data of five patients, as in Table 1, using binary values to indicate whether these symptoms exist or not in each case. The input variables are "Fever" and "Vomiting". The output variable is "Death". For example, for the first patient P_1 it is observed that he/she is vomiting and finally he/she dies; for the second patient P_2 it is observed that he/she has fever and finally he/she dies,…, etc. Therefore, what meaningful conclusions can be derived from these cases by the affinity model? First, we denote a rule by a triple $r = (x_1, x_2, y)$, then use Procedure 2.1:

1) Define the metric space (V, d). Define the referential set V as the set of all guesses/rules that can be used to identify the disease. Distance d is the failure (inaccurate prediction) rate of a rule (a distance concept), defined as the failure frequency of rule; d' is used to present the hit rate of the rule and $d' = 1 - \alpha d$. The hit rate is defined as the frequency of accurate prediction divided by the number of samples observed. According to Definition 2.13, d' is used to measure the degree of affinity of rules.

2) Determine the referential set X. The referential set X $=\{r_i,\ i=\overline{1,m}\}$, is a subset of V, the set of all possible rules/guesses completing the vector space to three dimensions. All the attributes are binary as shown in Table 1, i.e., $r = (x_1, x_2, y) \in X$, $x_1 \in \{0,1\}$, $x_2 \in \{0,1\}$ and $y \in \{0,1\}$. Because we use binary values here for attributes, only eight combinations/guesses can be generated with respect to three discrete attributes. Each rule $r_i \in X$, $i=\overline{1,8}$ competes for better affinity with respect to affinity set A, which is the set of rules capable of predicting the consequence of disease at the fixed time.

3) Choose subset B of X as the core of affinity set A. We choose B as the set containing the rules with the maximal hit rate.

4) Use affinity d' as defined:

$$d': X \to [0,1]$$
$$e \to d'(e, B) = 1 - \alpha d(e, B)$$

Finally, compute the hit rate (degree of affinity) of each rule in X, and select k for the k-Core (A). Because guesses/rules are limited to eight combinations, by simultaneously considering three attributes, we summarize the degree of affinity for each rule (r_i) as follows:

r_1 : if $x_1 = 1$ and $x_2 = 1$, then $y = 1$, miss rate = 5/5 , hit rate (affinity degree) = 1 – 5/5 = 0
r_2 : if $x_1 = 1$ and $x_2 = 1$, then $y = 0$, miss rate = 5/5, hit rate (affinity degree) = 1 – 5/5 = 0
r_3 : if $x_1 = 1$ and $x_2 = 0$, then $y = 1$, miss rate = 4/5, hit rate (affinity degree) = 1 – 4/5 = 1/5
r_4 : if $x_1 = 1$ and $x_2 = 0$, then $y = 0$, miss rate = 3/5, hit rate (affinity degree) = 1 – 3/5 = 2/5
r_5 : if $x_1 = 0$ and $x_2 = 1$, then $y = 1$, miss rate = 3/5, hit rate (affinity degree) = 1 – 3/5 = 2/5
r_6 : if $x_1 = 0$ and $x_2 = 1$, then $y = 0$, miss rate = 5/5, hit rate (affinity degree) = 1 – 5/5 = 0
r_7 : if $x_1 = 0$ and $x_2 = 0$, then $y = 1$, miss rate = 5/5, hit rate (affinity degree) = 1 – 5/5 = 0
r_8 : if $x_1 = 0$ and $x_2 = 0$, then $y = 0$, miss rate = 5/5, hit rate (affinity degree) = 1 – 5/5 = 0

After computation, we obtain the 0.2-core(A) = $\{r_3, r_4, r_5\}$; if $k = 0.4$, then the 0.4-core(A) = $\{r_4, r_5\}$. If a rule/guess, for instance, $r = (x_1, x_2, y)$ (or r_i) is capable of hitting the observed samples with a higher frequency (i.e., lower frequency of missing), then $r = (x_1, x_2, y)$ or r_i, has a greater degree of affinity with A, or rule r_i is useful/valuable to explain the behavior of the samples collected/observed. Thus, if we set $k = 0.4$, we can easily determine the 0.4-core(A) by two rules: Rule 4 tells that the $x_1 = 1$ (Fever) is not fatal, but Rule 5 warns the doctors that the $x_2 = 1$ (Vomiting) caused by this new disease could kill a patient. Of course, as the sample size increases, and as the variety of these qualitative attributes increases, using such simple thinking can approximate any affinity set A.

Readers may be confused about the difference between our affinity data-mining model and the model of association rules (Brossette et al., 1998); however, these two models are significantly different because: (a) a model of asso-

ciation rules uses the support and confidence of conditional probability to mine useful rules, but an affinity model uses the subjectively defined closeness occurrence frequency of rules; (b) an affinity model assumes that, for instance, $r = (x_1, x_2, y)$ is a vector in a metric/vector space, but the model of association rules does not make this assumption, and, more importantly, (c) it is possible to use various definitions in an affinity model in order to measure the degree of affinity. In this manner, it is not only possible, but easy to define the closeness between any two rules, or the distance from a rule to a specified group/set for further use without statistical restrictions.

3 Popular Data Mining Models

In this section, we present a brief review of several data mining models popularly used in medicine. These models include neural network (NN), rough set (Rosetta), support vector machine (SVM), decision tree (DT), association rule (AR) and logistic regression (LR). The LR model is popularly used in traditional statistical analysis in medicine (Delen, Walker, Kadam, 2005; Lavarc, 1999).

The amount of data collected and stored in medical databases has dramatically increased, due to advancements in automated data collection, and traditional data analysis techniques are no longer adequate for this volume of data (Brossette et al., 1998; Burke et al., 1997). For this reason, a number of non-traditional techniques have been developed to represent these values. For example, Delen et al. (2005) used artificial neural networks (ANN), decision trees (DT) and logistic regression (LR) to predict the survivability of breast cancer, concluding that ANN and DT both performed better than LR. Chang and Chen (2009) also used DT in combination with NN for skin diseases with prediction accuracy as high as 92.62%, which also outperformed LR. The rough set is another powerful model in this field (Pawlak, 1991). Wilk et al. (2005) described a rough methodology used for identifying the most relevant clinical features and for generating decision rules based on selected attributes from a medical data set with missing values. These rules could help (ER) medical personnel in the triage (initial assessment) of children with abdominal pain. Hirono and Tsumoto (2005) introduced a rough representation of a region of interest (ROI) in medical images. The main advantage of this method was its ability to represent inconsistencies between the knowledge-driven shape and image-driven shape of an ROI. As for the SVM, Meyfroidt et al. (2009) proposed a general overview of machine learning techniques, with a more detailed discussion of a number of these techniques to encouraging doctors to use them. They also provided guidance for applications and directions of research for SVMs. When using SVM to predict the depth of infiltration in endometrial carcinoma based on transvaginal sonography

(Spackman, 1991), SVMs were more effective than logistic regression. Bazzani et al. (2001) used an SVM classifier to distinguish false signals from microcalcifications in digital mammograms. The SVM classifier performed slightly better than a classifier implemented using an ANN. Van Gestel et al. (2004) compared least squares SVMs with DT, Naive Bayes, and LR for the classification of 20 benchmark datasets. They reported that SVMs exceeded the other methods in most of the datasets and were not significantly worse in the remaining datasets.

Decision trees (DTs) and association rules (ARs) are other valuable tools in medical data mining. For example, Mugambi et al. (2004), addressed this issue using a novel hybrid multivariate decision tree comprising polynomial, fuzzy and decision tree structures. As for the association rules method, Delgado et al. (2001), introduced a new fuzzy approach to association rules among quantitative values in relational databases. These fuzzy association rules were more informative than rules related to precise values. They also introduced a new means to measure accuracy, and claimed that their work was more understandable and appropriate than typical systems. Kuo and Shih (2007) applied an ant colony system (ACS) to a large database of health insurance to derive association rules, and showed that the newly proposed method was able to provide more condensed rules than an *a priori* method. Computation time was also reduced. In addition, the LR model is commonly used in medicine; for example, Spackman (1991), Tu (1996) and Doig et al. (1993) all used LR models in their studies. However, the performance of LR was inferior to that of NN models.

To summarize, the above data mining models made considerable contributions to overcoming the problems associated with data mining. We simply compare the aforementioned models as in Table 2 for their advantages and disadvantages.

Table 2: Comparison of Data Mining Models

Characteristics/Models	SVM	NN	DT	LR	AR
Advantages	The prediction power is very strong	The graphical construction of model is clear	It is easy to use and explain	It is easy to use and explain	It is easy to catch the relationship between causes and consequences
Disadvantages	It is difficult to describe the clear rules between causes and consequences	It is difficult to describe the clear rules between causes and consequences	It is difficult to group and cluster when data are huge	The explanatory power is weak if the data do not follow the statistical assumptions	The explanatory power is weak if the data do not follow the statistical assumptions

Next, we compare the performance of the affinity model with that of the aforementioned models. The challenge for all of the data mining models is in the fact that the sample size was not large (only 645 units), and no statistical distribution was pre-assumed for the data.

4 Multi-objective Affinity Model for Data Mining

In this study, Step 4 in Procedure 2.1 was extended to consider multi-objectives of affinity. In Procedure 2.1, it was logical and reasonable for the decision maker to select the value of k first; for example, Michnik et al. (2008) proposed a similar idea using the iterated algorithm to find the final k-core(A). However, it was not easy to operate in this manner for most actual cases, and selecting the value of k at the beginning is a particularly difficult task for inexperienced decision makers. Early in 2006, Wu et al. (2009) used a multi-objective affinity classification system comparing ant colony optimization (ACO) in the classification of delayed diagnostics, and concluded that the multi-objective affinity set classification system was superior to the ACO system. Their fitness function of two objectives: z_1, z_2 is as follows (Wu et al., 2009):

$$f(z_1, z_2) = w_1 \times (N - z_1) + w_2 \times z_2 \tag{5}$$

where:

z_1 − number of rules in a subset, $z_1 < N$;

z_2 − prediction accuracy of rules in a subset;

N − maximal number of rules in a subset predetermined by the decision maker;

w − weight of objective predetermined by the decision maker.

In the above paper, Wu et al. (2009) used the weighing objective function (5) to rank the appropriate subset of rules by setting $w_1 = w_2 = 0.5$. Because z_1 and z_2 were not in the same scale, the performance of z_1 could be over-emphasized. In addition, Chen et al. (2009) used multi-objective ideas rather than selecting the value of k, and separated the data set into a training set and validation set, proposing two criteria to select the final k-core(A): one was that each rule had to include at least two causes (x), the other was that the rule base had to be able to catch the validation set 100% of the time. Thus, $\mathrm{M}_{\mathrm{A}}^{e}(w^0) \geq 0.247$ or $k = 0.247$ were finally achieved.

The study of Wu et al. (2009) did not demonstrate the potential power of multi-objective affinity classification system, which inspired us to compare the multi-objective affinity model with many traditional data mining methods. Furthermore, our fitness function for ranking the subset of rules was based on affinity, on which values ranged from 0 to 1 (normalized). This study extended and modified the research of Chen et al. (2009) and Wu et al. (2009) to a multi-objective problem (Steuer, 1986). We assumed that a decision maker is unable to

select the value of k in the beginning, but has multiple goals to form the k-core (A). For example, he/she may want to minimize the size of the k-core(A), i.e., the number of rules is decreased, but desires the prediction accuracy of the k-core(A) to remain high. In such situations, there are conflicts between two goals, in attempting to minimize the number of rules while maximizing the prediction accuracy of the rule base. Each rule set presents a possible feasible solution, and each rule set plays the role of set B in Procedure 2.1. In this case, B is evaluated by its objective of minimizing the number of rules and simultaneously maximizing the prediction accuracy. In Section 3, these two objectives are clearly defined according to their affinities. To achieve this, the affinity d' in Step 4 of Procedure 2.1 is newly defined by integrating the affinities of the aforementioned two objectives.

The following is used to illustrate our new multi-objective approach to computing d' in step 4 of Procedure 2.1. First, an initial rule set C of the best 100 rules with highest affinities is prepared by Procedure 2.1. Here, too, we use the idea of Example 2.1. If rule r_i is found in the training set once, then its corresponding affinity degree is one divided by the size of the training set; if rule r_i is found in the training set twice, then its corresponding affinity degree is two divided by the size of training set, and so on. The degree of affinity for a rule in the training set is used as the prediction reference for the validation set, which is denoted by aff_{r_i} in the following. It is logical to say that if a rule is frequently found in the training set, then it has a higher degree of prediction power for the validation set and should be kept in C. Second, assume set B is randomly generated and B$\subseteq$C. B is chosen to approximate the final core of affinity set A. If the size of B, i.e., the number of rules in B, is *norm*(B), then our first affinity d'_1 is defined as follows:

$$d'_1 = \min_{r_i \in \mathrm{B}} [\frac{aff_{r_i}}{norm(\mathrm{B})}] \tag{6}$$

Third, we assume that the decision maker expects the number of rules in the final core to be small, but he hopes that it will contain at least fifteen rules. When the number of rules is more than fifteen, his satisfaction is reduced. Thus, we can simply define the second affinity d'_2 as follows:

$$d'_2 = \frac{15}{norm(\mathrm{B})} \tag{7}$$

Here *norm*(B) is the size of B and $15 \le norm(\mathrm{B}) \le 30$ is assumed in this study. Thus, the new d' is defined as the well-known weighted function in multi-objective programming theory (Steuer, 1986):

$$d' = w_1 d'_1 + w_2 d'_2 \tag{8}$$

where $w_1 + w_2 = 1$ and $w_1, w_2 \geq 0$. The weights of (w_1, w_2) are selected subjectively at the beginning. According to the new definitions above and Procedure 2.1, the iteration steps of this study are as follows:

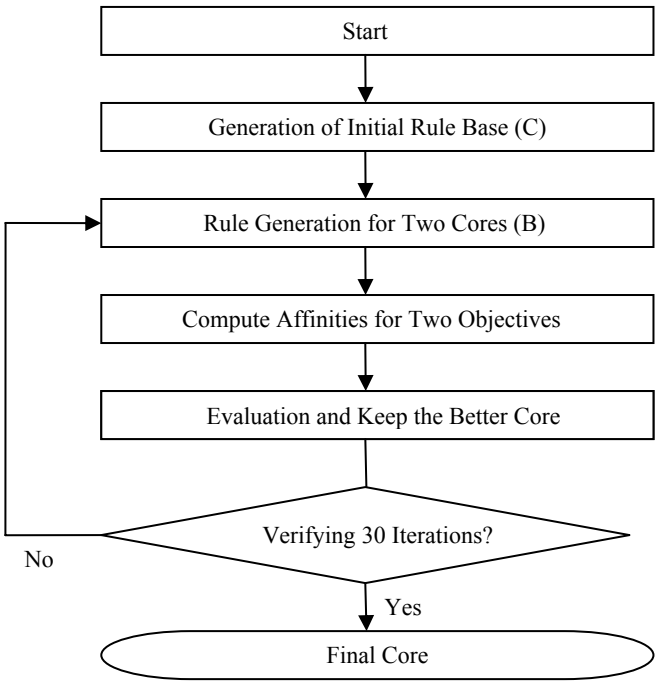

Figure 2. Process of Data Mining using the Multi-objective Affinity Model

Step 0. Subjectively set the pair (w_1, w_2). In this study, w_2 is set to 0.6 and w_1 is set to 0.4. This means we emphasize fewer rules to catch more observations. This is the *Start* stage.

Step 1. Separate the sample data into two parts; for example, 80% of data are used for training and 20% for validation. At the same time, aff_{r_i} for each rule r_i is computed in this stage and Procedure 2.1 is followed exactly to implement this step. We set a threshold to generate the initial rule base C: although thousands of rules are generated by Procedure 2.1, only the rules with the top 100 affinities are retained. This is the stage of *Generation of Initial Rule Base*.

Step 2. Randomly generate two rule sets, for instance, B_1, $B_2 \subseteq C$, to approximate the core (A). Each rule r_i in B_j, $j = 1, 2$ has its causal part (x) and consequence part (y). The size of B_j, i.e, $norm(B_j)$ is also different for each rule set, but it is included between 10 and 30. Only two cores are generated at the beginning. This is the stage of *Rule Generation for Two Cores*.

Step 3. Apply Equation (4) to compute the minimal degree of affinity d_1' for each B_j, and apply Equation (5) to compute d_2': the satisfaction felt by the decision maker with the size of B_j. After that, $d' = w_1 d_1' + w_2 d_2'$ defined in Equation (8) is used to evaluate each B_j. In this case, B_j, j = 1, 2, subsets of X, are chosen as candidates for being the core of affinity set A (where core(A) is that set B for which d' = 1). This is the stage of *Computing Affinities for Two Objectives*.

Step 4. Keep only that B_j for which *d'* is largest in Step 3 and return to ***Step 2*** to generate another B. This is the stage of *Evaluation and Keeping the Better Core*.

Step 5. Repeat the steps 1–4 until the predetermined number of iterations has been reached. Here the number of iterations is set to 30. This is the stage of *Verifying 30 Iteration*s.

Step 6. If 30 iterations are reached in ***Step 5***, then output B as the approximated core of A. This is the stage of *Final Core*.

Using these steps, 645 samples were used for training the neural network, rough set model (Rosetta), supporting vector machine (SVM), decision tree, association rule, logistic regression and the multi-objective affinity model: the performance of these models is compared in Section 5.

5 Actual Example

The objective of this research was to identify the core attributes leading to frequent revisits of emergency patients in ED within a set period of time; simply speaking, doctors expect generating useful rules for avoiding revisits. The study uses the original data from the website of Kaoping Area Medical Emergency Response Alliance (KAMERA). This site is the largest site in Taiwan for collecting trauma data of patients by more than 30 hospitals joining in an alliance. Doctors presented 645 samples of clinical data from 2008 (from Jan. to Dec.), and the samples were divided into two parts: the training set and the validation set. The training-validation ratio was established as 80%-20%, 70%-30% and 60%-40% of the data. The training set was used to derive rules from various data mining models and the validation set was used to draw the ROC curve to compare the performance of each model. On the basis of the availability of data retrieved from electronic medical records, physicians suggested nine possible influential attributes/causes $\{x\}$ leading to emergency patient revisits of (y); age (x_1), triage status (x_2), healthcare provider (x_3), time of visit (x_4), length of ED stay (x_5), breathing pattern (x_6), blood-pressure (x_7), pulse rate (x_8), physician-patient ratio, (x_9) and revisiting frequency (y), as shown in Table 3. The physician-patient ratio was defined as the number of on-duty physicians divided by the number of the patients in the ED within an 8-hour shift.

Table 3: Attributes of the Data Mining Model

Attributes	Interval	Coding
Age (x_1)	0-8	1
	9-18	2
	19-40	3
	41-65	4
	Over 66	5
Triage status (x_2)	Level 1 (Severe)	1
	Level 2 (Moderate)	2
	Level 3 (Mild)	3
Healthcare provider (x_3)	Pediatric emergency	1
	Emergency medicine	2
	Surgical emergency	3
	Others	4
Time of visiting (x_4)	00:00-08:00	1
	08:00-16:00	2
	16:00-24:00	3
Length of ED stay (x_5)	0-4 hours	1
	4-8 hours	2
	8-12 hours	3
	Over 12 hours	4
Breath pattern (x_6)	Normal	1
	Abnormal	2
Blood-pressure (x_7)	Normal	1
	Abnormal	2
Pulse rate (x_8)	Normal	1
	Abnormal	2
Physician – patient ratio (x_9)	High (1~1/20)	1
	Moderate (1/20~1/40)	2
	Low (Under 1/40)	3
Revisiting frequency (y)	One time	0
	More than one time	1

Note: the index of medical capacity is defined as the number of the available doctors divided by the number of the patients in ED.

The referential set X is defined as the vector space with the dimensionality of ten and attributes are discrete as in Table 3, $r = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, y) \in X$ by Definition 2.10. The value of each x_i (i = 1,2, … ,9) and y were randomly selected from the attribute domain in Table 3. If any x_i (i = 1,2, … ,9) had a value of zero, then this means that the corresponding attribute x_i would not be considered in the formation of rules.

Here, our new model and the popular data mining models above will be tested for their performance using the confusion matrix and ROC curve.

5.1 Confusion Matrix and ROC Curve

We employed the confusion matrix (Collett, 2003) to compare the performance of our multi-objective affinity model and of other popular data mining models. In artificial intelligence, particularly for the binary consequences of information systems, a confusion matrix is a visualization tool typically used in supervised learning. Each column of the matrix represents instances in a predicted class, while each row represents instances in an actual class. One benefit of a confusion matrix is that it is easy to observe whether the system is confusing two classes (i.e. commonly mislabeling one as another). For example, the following Table 4 shows the confusion matrix for a two-class classifier. The entries in the confusion matrix have the following meaning in the context of our study: *a* is the number of correct predictions that an instance is negative, *b* is the number of incorrect predictions that an instance is positive, *c* is the number of incorrect predictions that an instance is negative, and *d* is the number of correct predictions that an instance is positive (Collett, 2003).

Table 4: Confusion Matrix

		Predicted	
		Negative	Positive
Actual	Negative	*a*	*b*
	Positive	*c*	*d*

Several standard terms should be defined for this matrix:

- *Accuracy* (*AC*) is the proportion of the total number of predictions that were correct. It is determined using the equation:

$$AC = \frac{a+d}{a+b+c+d}$$

- The *recall* or *true positive rate* (*TP*) is the proportion of positive cases that were correctly identified, as calculated using the equation:

$$TP = \frac{d}{c+d}$$

- The *false positive rate* (*FP*) is the proportion of negatives cases that were incorrectly classified as positive, as calculated using the equation:

$$FP = \frac{b}{a+b}$$

- The *true negative rate* (*TN*) is defined as the proportion of negatives cases that were classified correctly, as calculated using the equation:

$$TN = \frac{a}{a+b}$$

- The *false negative rate* (*FN*) is the proportion of positives cases that were incorrectly classified as negative, as calculated using the equation:

$$FN = \frac{c}{c+d}$$

- Finally, *precision* (*P*) is the proportion of the predicted positive cases that were correct, as calculated using the equation:

$$P = \frac{d}{b+d}$$

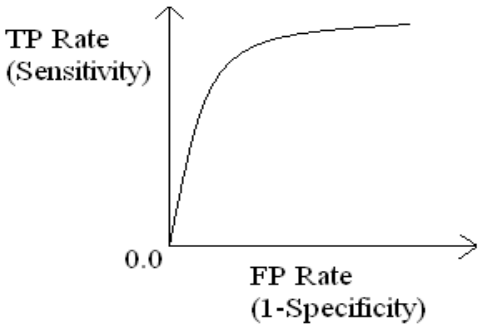

Figure 3. ROC Curve

In addition, once the confusion matrix was prepared, the ROC curve could be easily drawn. The receiver operating characteristic (ROC) curve (Zweig, Campbell, 1993) was used to compare the performance of our affinity model and of other models. In signal detection theory, a receiver operating characteristic (ROC), or simply ROC curve, is a plot of the sensitivity vs. 1 – specificity for a binary classifier system as its discrimination threshold is varied. The ROC can also be represented in the form of TP (true positive), FP (false positive), TN (true negative) and FN (false negative). For example, if a rule predicts that a patient has a high frequency of revisits (positive), and it really happens, then this is a TP case; on the contrary, if it doesn't happen then this is an FP case. The number of TPs and TNs should be reasonably large for a good prediction model. The diagnostic performance of a test or the accuracy of a test to distinguish cases of disease from normal cases was evaluated using ROC curve analysis (Zweig, Campbell, 1993). Receiver operating characteristic (ROC) curves can also be

used to compare the diagnostic performance of two or more laboratory or diagnostic tests (Collett, 2003) − see Figure 3. If the plotted ROC curve of a model is more north-west skewed, or the area under the ROC curve is larger, then this model is more beneficial. The confusion matrices and the ROC curves are available in Section 4 for each data mining model.

5.2 Performance of Models

Case I, Case II and Case III show the results of training-validation rates at 80%--20%, 70%-30%, and 60%-40%, respectively. For simplicity, in the following tables, we use MA for the multi-objective affinity model, NN for the neural network model, RS for the rough set model, SVM for the model of supporting vector machine, DT for the decision tree model, AR for the model of association rules and LR for logistic regression model. For the accuracy and TP indeces the larger value the better; while for the FP index the converse is true: the smaller value the better. The ROC curve was used to compare these models in the end.

Case I: Training-validation rate of 80%-20%
The performance of each model for Case I is summarized in the following Tables 5-6.

Table 5: Confusion Matrix of Case I

Actual/Predicted	0	1
0	55(MA), 44(NN), 41 (RS), 62(SVM), 32(DT), 28(AR), 62(LR)	12(MA), 24(NN), 27(RS), 6(SVM), 36(DT), 40(AR), 6(LR)
1	13(MA), 18(NN), 24(RS), 17(SVM), 30(DT), 24(AR), 53(LR)	50(MA), 43(NN), 37(RS), 44(SVM), 31(DT), 37(AR), 8(LR)

Table 6: Performance of Case I

Model	MA	NN	RS	SVM	DT	AR	LR
Accuracy	81.6%	67.4%	60.5%	82.2%	48.8%	50.4%	54.3%
TP	78.7%	70.5%	60.7%	72.1%	50.8%	60.7%	13.1%
FP	17.6%	30.3%	39.7%	8.8%	52.9%	58.8%	8.8%

In the first case, SVM performed best (Accuracy = 82.2%), MA was a little behind SVM (Accuracy = 81.6%). In addition, the decision tree model had the poorest performance (Accuracy = 48.8%).

Case II: Training-validation rate of 70%-30%
The performance of each model for Case II is summarized in the following Tables 7-8.

Table 7: Confusion Matrix of Case II

Actual/Predicted	0	1
0	78(MA), 63(NN), 57(RS), 88(SVM), 45(DT), 43(AR), 3(LR)	16(MA), 31(NN), 37(RS), 6(SVM), 49(DT), 51(AR), 91(LR)
1	20(MA), 34(NN), 40(RS), 26(SVM), 45(DT), 51(AR), 0(LR)	81(MA), 66(NN), 60(RS), 74(SVM), 55(DT), 49(AR), 100(LR)

Table 8: Performance of Case II

Model	MA	NN	RS	SVM	DT	AR	LR
Accuracy	81.6%	66.5%	60.3%	83.5%	51.5%	47.4%	53.1%
TP	79.0%	66.0%	60.0%	74.0%	55.0%	49.0%	100%
FP	16.0%	33.1%	39.4%	6.4%	52.1%	54.3%	96.8%

In the second case, SVM performed best (Accuracy = 83.5%), and MA was still a little behind SVM (Accuracy = 81.6%). In this case, the model of association rules had the lowest accuracy of 47.4%. Furthermore, logistic regression had had uncommonly high TP and FP, which hints that the performance of this model is unstable.

Case III: Training-validation rate of 60%-40%
The performance of each model for Case III is summarized in the following Tables 9-10.

Table 9: Confusion Matrix of Case III

Actual/Predicted	0	1
0	103(MA), 80(NN), 77(RS), 119(SVM), 68(DT), 64(AR), 48(LR)	26(MA), 47(NN), 50(RS), 8(SVM), 59(DT), 63(AR), 79(LR)
1	25(MA), 43(NN), 52(RS), 33(SVM), 51(DT), 62(AR), 48(LR)	106(MA), 88(NN), 79(RS), 98(SVM), 80(DT), 69(AR), 83(LR)

Table 10: Performance of Case III

Model	MA	NN	RS	SVM	DT	AR	LR
Accuracy	82.2%	65.1%	60.5%	84.1%	57.4%	51.6%	50.7%
TP	79.9%	67.2%	60.3%	74.8%	61.1%	52.7%	63.3%
FP	21.0%	37.0%	39.4%	6.3%	53.5%	49.6%	62.2%

In the third case, SVM performed best (Accuracy = 84.1%), followed by MA (Accuracy = 82.2%). Moreover, logistic regression had the poorest accuracy of 50.7%. Finally, the ROC curves and the area under each model are presented in the following, to illustrate the computational results above.

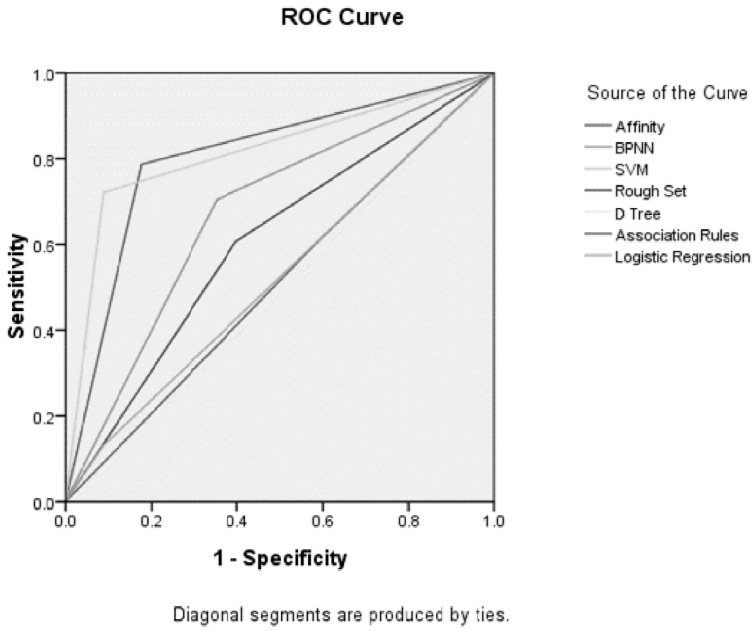

Figure 4. ROC Curve for Training-validation Ratio of 80%-20%

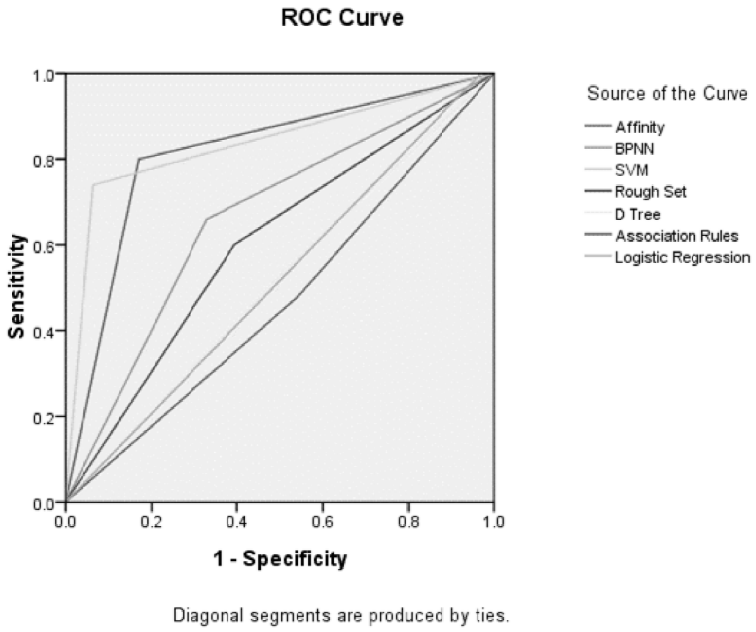

Figure 5. ROC Curve for Training-validation Ratio of 70%-30%

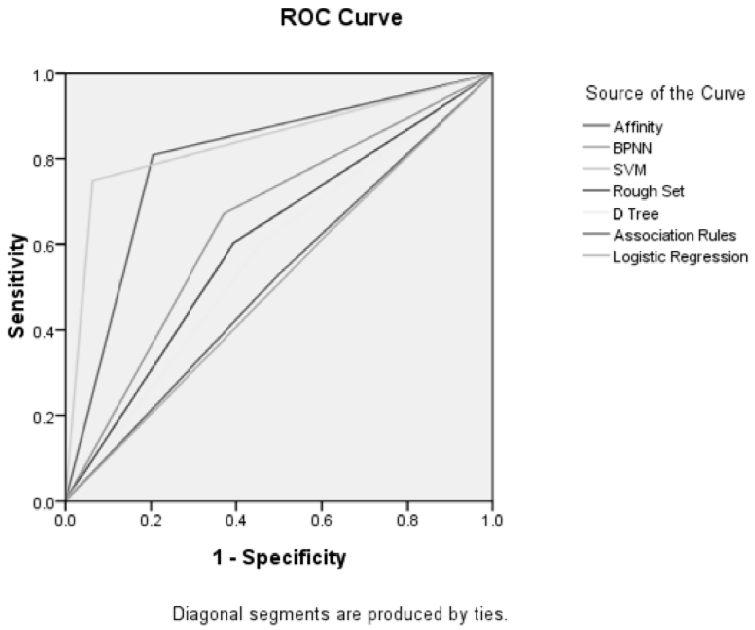

Figure 6. ROC Curve for Training-validation Ratio of 60%-40%

Table 11: Area under each Model of ROC Curves

Model	MA	NN	RS	SVM	DT	AR	LR
Case I	0.810	0.616	0.605	0.817	0.489	0.509	0.521
Case II	0.820	0.665	0.603	0.838	0.514	0.469	0.516
Case III	0.812	0.651	0.605	0.843	0.573	0.515	0.506

According to Table 11, if the area under the curve were larger, then it would have a better classification power. We simply concluded that: SVM $\succ$ MA $\succ$ NN $\succ$ RS, where "$\succ$" means "to be superior to". To summarize, these tables and ROC curves show that the multi-objective affinity model and the SVM model had significant advantages over the other data mining models. Although SVM had the best classification power, considering that the objective of this study was to find rules, we continued using the multi-objective affinity model for further explanations.

The multi-objective affinity model generated seventeen rules with a compromised d' of 0.3757. These rules are summarized in the following table, illustrating the causes leading to a high frequency of revisits.

Table 12: Generated Rules of the Multi-objective Affinity Model

Rule	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	y
r_1	-	3	-	-	-	1	1	-	-	1
r_2	-	-	-	-	1	1	1	-	-	1
r_3	-	3	-	-	1	-	-	-	-	0
r_4	-	2	-	-	-	-	-	-	2	1
r_5	-	3	2	-	-	-	-	-	-	0
r_6	-	-	-	-	1	1	-	-	2	1
r_7	-	3	-	-	-	-	-	-	2	1
r_8	-	-	2	-	1	-	-	-	-	0
r_9	-	-	-	-	-	1	1	-	2	1
r_{10}	-	-	2	-	-	1	1	-	-	1
r_{11}	-	-	2	-	-	-	2	-	-	0
r_{12}	-	3	-	-	1	-	1	-	-	1
r_{13}	-	-	-	3	1	-	-	-	-	0
r_{14}	-	3	2	-	-	-	-	-	2	1
r_{15}	-	3	-	-	-	-	2	-	-	0
r_{16}	-	3	-	-	-	-	1	-	2	1
r_{17}	-	3	-	3	-	-	-	-	-	0

Note: "-" means that the corresponding attribute is ignored.

According to Table 12 and the definition of variables in Table 3, we focus on the causes $\{x_i\}$, which lead y to 1. Here x_2 ranges from 2 to 3, x_3 is at most 2, x_5 is at most 1, x_6 is at most 1, x_7 is at most 1 and x_9 is at most 2. Therefore, these rules (grey squares) could be interpreted as follows: if a patient's triage scale (x_2) is two or three, or visiting service (x_3) is in the division of emergency medicine, or stay in the ED (x_5) is less than four hours, or breath pattern (x_6) appears normal, or blood pressure (x_7) is within normal limit, or the physician-patient ratio (x_9) is in the middle level, then the revisiting frequency (y) is high. Interestingly, the mining results of $\{x_6, x_7\}$ above closely match the conclusions in Chen et al. (2009). That is, when the patient looks fine, then his/her frequency of revisiting the ED could be high.

5.3 Discussions

The following discussions are results of brain storming with the physicians using their clinical experiences. According to the results of this study, patients with abnormal blood pressure and breath patterns revisited less frequently. It is commonplace for physicians to pay more attention to patients with unstable vital signs (Aaland, Smith, 1996; Chen et al., 2009) rather than to those patients who appear normal. In such cases, more real-time, comprehensive, continuous and thorough/whole examinations tend to be performed and developed, and their

problems are more likely to be addressed adequately during their initial stay in the ED, thereby avoiding possible revisits. On the other hand, the patients triaged as levels 2 or 3 are conventionally termed non-critical patients.

Our results show that a physician-patient ratio at a moderate level is associated with a higher rate of revisits. This could result from the fact that when a physician cares for too many patients, he/she will fail to provide adequate medical service for all of them. Nevertheless, a higher revisit rate was not found in the group with low physician-patient ratio.

To summarize, we propose the following issues:

1) Compared to the level 1 group in triage, groups 2 and 3 are relatively ambulatory, with less severity of illnesses. They might receive less medical treatment with fewer aggressive interventions, resulting in more unplanned revisits. The aforementioned observation tells us that the patient's situation in ED is dynamic and unpredictable, and therefore an innovative, complete and effective process for examining patients is required.
2) The low physician-patient ratio could impair the operational efficiency of the ED, thereby blocking patient's intention to revisit. Having the impression of receiving suboptimal care in the same ED, those patients may seek aid in other hospitals. However, this assumption needs more evidences to prove.
3) Humans are fallible, also in their observations of patients. If the medical personnel (doctors and nurses) is not able to pay full attention to patients in the short run, then a real-time and whole process for examining the vital signs of patients is suggested. Therefore, wearable devices for ED patients could be valuable. We could respond faster and more correctly by continuously monitoring or early alerting these patients to avoid unplanned revisits.

6 Conclusions and Recommendations

The explanatory power of the affinity model is better than that of most of the existing models. However, the data collected in this study regarding revisiting patients may have lacked some important/hidden attributes/features, detracting from the effectiveness of the mining results. The affinity model will certainly be able to provide decision makers with more satisfactory results, once the structural model is further enhanced. Other mapping/projection methods based on affinity may also generate effective rules to overcome problems associated with data mining. It is worth noting that: (a) the affinity model is quite simple, (b) it does not require explicit membership functions (Zadeh, 1965), and (c) it has significantly better performance than existing models. For further research, we propose the application of the affinity model to more complex data mining medical problems and other areas.

Acknowledgements

This research is funded by National Science Council of Taiwan (96-2416-H-212-002-MY2), and Kaohsiung Medical University (Kmuh-96-6G07, Kmuh-98-8G30). Authors would like to appreciate Prof. Chih-Hung Wu, Prof. Jerzy Michnik, Miss Chih-Min Shen and Mr. Zheng-Yen Hsieh for their serious efforts in this area.

References

Aaland M.O., Smith K. (1996), *Delayed Diagnosis in a Rural Trauma Centre*, Surgery 120, 774-9.

Abu-Hanna A., de Keizer N. (2003), *Integrating Classification Trees with Local Logistic Regression in Intensive Care Prognosis*, Artificial Intelligence Medicine, 29, 5-23.

Achenbach P., Koczwara K., Knopff A., Naserke H., Ziegler A.G., Bonifacio E. (2004), *Mature High-affinity Immune Responses to (Pro)insulin Anticipate the Autoimmune Cascade that Leads to Type 1 Diabetes*, Clin Invest., 114, 589-97.

Aguilar-Ruiz J.S., Costa R., Divina F. (2004), *Knowledge Discovery from Doctor-patient Relationship*, Proceedings of SAC '04, March 14-17, Nicosia, Cyprus.

Alanazi H.O., Abdullah A.H., Larbani M. (2013), *A Critical Review for Developing Affinity Set Method for Multi Classification and Prediction*, International Journal of Computer Science Engineering and Technology 3, 394-5.

Bazzani A., Bevilacqua A., Bollini D. et al. (2001), *An SVM Classifier to Separate False Signals from Microcalcifications in Digital Mammogram*, Physics in Medicine and Biology, 46, 1651-63.

Berman J.J. (2002), *Confidentiality Issues for Medical Data Miners*, Artificial Intelligence in Medicine 26, 25-36.

Brooksa A., Holroydb B., Riley B. (2004), *Missed Injury in Major Trauma Patients*, Injury, Int. J. Care Injured, 35, 407-10.

Brossette S.E., Sprague A.P., Hardin J.M., Waites K.B., Jones W.T., Moser S.A. (1998), *Association Rules and Data Mining in-hospital Infection-control and Public-health Surveillance*, Journal of the American Medical Informatics Association, 5, 373-81.

Burke H.B., Goodman P.H., Rosen D.B., Henson D.E., Weinstein J.N., Harrell Jr F.E. et al. (1997), *Artificial Neural Networks Improve the Accuracy of Cancer Survival Prediction*, Cancer, 79, 857-62.

Burke H.B., Rosen D.B., Goodman P.H. (1995), *Comparing the Prediction Accuracy of Artificial Neural Networks and Other Statistical Models for Breast Cancer Survival* [in:] G. Tesauro, D. Touretzky, T. Leen (eds.), *Advances in Neural Information Processing Systems*, Vol. 7, MIT Press, Cambridge, MA, 1063-7.

Cattell V. (2001), *Poor People, Poor Places, and Poor Health: The Mediating Role of Social Networks and Social Capital*, Social Science & Medicine, 52, 1501-1516.

Chen Y.W., Larbani M., Hsieh C.Y., Chen C.W. (2009), *Introduction of Affinity Set and Its Application in Data Mining Example of Delayed Diagnosis*, Expert Systems with Applications, 36, 10883-9.

Chang C.L., Chen C.H. (2009), *Applying Decision Tree and Neural Network to Increase Quality of Dermatologic Diagnosis*, Expert Systems with Applications, 36, 4035-41.

Collett D. (2003), *Modeling Binary Data*, Chapman & Hall/CRC.

Corne D., Dhaenens C., Jourdan L. (2012), *Synergies between Operations Research and Data Mining: The Emerging Use of Multi-objective Approaches*, European Journal of Operational Research, 221, 469-79.

Cortes C., Vapnik V. (1995), *Support-Vector Networks*, Machine Learning, 20, 273-97.

Delen D., Walker G., Kadam A. (2005), *Predicting Breast Cancer Survivability: A Comparison of Three Data Mining Methods*, Aritificial Intelligence in Medicine, 34, 113-27.

Delgado M., Sanchez D., Martin-Bautista M.J., Vila M.A. (2001), *Mining Association Rules with Improved Semantics in Medical Databases*, Artificial Intelligence in Medicine, 21, 241-5.

Doig G.S., Inman K.J., Sihhald W.J., Mattm C.M., Rohetrson J.M. (1993), *Modeling Mortality in the Intensive Care Unit: Comparing the Performance of a Hack-propagation, Associative-learning Neural Network with Multivariate Logistic Regression*, Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care, McGraw-Hill, Inc., Washington, DC, 361-5.

Esfandiaria N., Babavaliana M.R., Moghadama A.E., Tabarb V.K. (2014), *Knowledge Discovery in Medicine: Current Issue and Future Trend*, Expert Systems with Applications, 41, 4434-63.

Freitas A.A. (2004), *A Critical Review of Multi-Objective Optimization in Data Mining: A Position Paper*, SIGKDD Explorations, 6, 77-86.

Freitas A.A. (2008), *A Review of Evolutionary Algorithms for Data Mining* [in:] O. Maimon, L. Rokach (eds.), *Soft Computing for Knowledge Discovery and Data Mining*, Springer Verlag, 79-111.

Furnival R.A., Woodward G.A., Schunk J.E. (1996), *Delayed Diagnosis of Injury in Pediatric Trauma*, Pediatrics, 98, 56-62.

Gestel T. van, Suykens J.A.K., Baesens B., Viaene S., Vanthienen J., Dedene G., Moor B., Vandewalle J. (2004), *Benchmarking Least Squares Support Vector Machine Classifiers*, Journal of Machine Learning Research, 54, 5-32.

Gordon J.A., An L.C., Hayward R.A. et al. (1998), *Initial Emergency Department Diagnosis and Return Visits: Risk versus Perception*, Ann Emerg Med, 32, 569-73.

Grupe F.H., Owrang M.M. (1995), *Data Base Mining Discovery New Knowledge and Cooperative Advantage*, Information System Management, 12, 26-31.

Hanlon J.J., Pickette G.E. (1979), *Public Health Administration and Practice*, CV Mosby Co., 711.

Hirono S., Tsumoto S. (2005), *Rough Representation of a Region of Interest in Medical Images*, International Journal of Approximate Reasoning, 40, 23-34.

Hu S.C. (1992), *Analysis of Patients Revisits to the Emergency Department*, Am J Emerg Med, 10, 366-70.

Hunt J., Cooke D. (1995), *An Adaptive, Distributed Learning System, Based on Immune System*, Proceedings of the IEEE International Conference on Systems, Man and Cybernetics 1995, 2494-9.

Kohn L.T., Corrigan J.M., Donaldson M.S. (1999), *To Err Is Human: Building a Safer Health System*, Washington D.C.

Kuo R.J., Shih C.W. (2007), *Association Rule Mining through the Ant Colony System for National Health Insurance Research Database in Taiwan*, Computers and Mathematics with Applications, 54, 1303-18.

Larbani M., Chen Y.W. (2009), *A Fuzzy Framework for the Concept of Affinity*, Applied Mathematical Sciences, Vol. 3, No. 7, 297-316.

Lavarc N. (1999), *Selected Techniques for Data Mining in Medicine*, Artificial Intelligence in Medicine, 16, 3-23.

Leape L.L., Brennan T.A., Laird N., Lawthers A.G., Localio A.R., Barnes B.A., Hebert L., Newhouse J.P., Weiler P.C., Hiatt H. (1991), *Incidence of Adverse Events and Negligence in Hospitalized Patients. Results of the Harvard Medical Practice Study I*, The New England Journal of Medicine, 324, 370-6.

Marcantonio E.R., Sylvia McKean S.M., Goldfinger M., Kleefield S., Yurkofsky M., Brennan T.A. (1999), *Factors Associated with Unplanned Hospital Readmission among Patients 65 Years of Age and Older in a Medicare Managed Care Plan*, The American Journal of Medicine, 107(1), 13-17.

Matejtshuk P. (eds.) (1997), *Affinity Separation: A Practical Approach*, Oxford University Press.

Meyfroidt G., Güiza F., Ramon J., Bruynooghe M. (2009), *Machine Learning Techniques to Examine Large Patient Databases*, Best Practice & Research Clinical Anesthesiology, 23, 127-43.

Michnik J., Michnik A., Pietuch B. (2008), *Credit Scoring Model Based on the Affinity Set*, Proceedings of the 10th International Conference on Enterprise Information Systems, 14-16, June, Barcelona, Spain.

Mugambi E.M., Hunter A., Oatley G., Kennedy L. (2004), *Polynomial-fuzzy Decision Tree Structures for Classifying Medical Data*, Knowledge-Based Systems, 17, 81-7.

Nuñez S., Hexdall A., Aguirre-Jaime A. (2006), *Unscheduled Returns to the Emergency Department: An Outcome of Medical Errors*? Qual Saf Health Care, 15, 102-8.

Paoin W. (2011), *Lessons Learned from Data Mining of WHO Mortality Database*, Methods of Information in Medicine, 50, 380-5.

Pawlak Z. (1991), *Rough Sets. Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers.

Pierce J.M., Kellerman A.L., Oster C. (1990), *Bounces: An Analysis of Short-term Return Visits to a Public Hospital Emergency Department*, Ann Emerg Med, 19, 752-7.

Spackman K.A. (1991), *Maximum Likelihood Training of Connectionist Models: Comparison with Least Squares Hack-propagation and Logistic Regression*, Proc Annu Symp Comput Appl Med Care, 285-9.

Steuer R.E. (1986), *Multiple Criteria Optimization: Theory, Computation and Applications*, John Wiley and Sons, New York.

Tu J.V. (1996), *Advantages and Disadvantages of Using Artificial Neural Networks versus Logistic Regression for Predicting Medical Outcomes*, J Clin Epidemid, 49, 1225-31.

Ve-McConnell (1999), *Who Is My Neighbor? Social Affinity in a Modern World*, State University of New York Press.

Wei G.L., Liu X.L., Li J.H., Liu Y., Shang Z.H. (2002), *Endotoxin Removal in Some Medicines and Human Serum Albumin Solution by Affinity Membrane*, Chinese Journal of Chromatography, 23, 108-14.

Wu C.H., Li W.T., Hsu C.C., Li C.H., Fang C.I., Wu C.H. (2009), *A Novel Multi-objective Affinity Set Classification System: An Investigation of Delayed Diagnosis Detection*, Proceeding of Intelligent Information and Database Systems, 1-3, 289-94.

Wu C.L., Wang F.T., Chiang Y.C., Chiu Y.F., Lin T.G., Fu L.F., Tsai T.L. (2008), *Unplanned Emergency Department Revisits within 72 Hours to a Secondary Teaching Referral Hospital in Taiwan*, The Journal of Emergency Medicine, 38, 512-517.

Wilk S.Z., Słowinski R., Michałowski W., Greco S. (2005), *Supporting Triage of Children with Abdominal Pain in the Emergency Room*, European Journal of Operational Research, 160, 696-709.

Zadeh L.A. (1965), *Fuzzy Sets*, Information and Control, 8, 338-53.

Zbikowski R., Hunt K.J. (ed.) (1996), Neural Adaptive Control Technology, World Scientific.

Zinkhan G.M. (2002), *Relationship Marketing: Theory and Implementation*, Journal of Market-Focused Management, 5, 83-9.

Zweig M.H., Campbell G. (1993), *Receiver-operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine*, Clinical Chemistry, 39, 561-77.