Expert Systems with Applications 36 (2009) 10883–10889

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Introduction of affinity set and its application in data-mining example of delayed diagnosis

Yuh-Wen Chen [a,*], Moussa Larbani [b], Cheng-Yen Hsieh [a], Chao-Wen Chen [c]

[a] Institute of Industrial Engineering and Management of Technology, Da-Yeh University, 112 Shan-Jeau Rd., Da-Tsuen, Chang-Hwa 51505, Taiwan
[b] Department of Business Administration, Kulliyyah of Economics and Management Sciences, IIUM University, Jalan Gombak, 53100 Kuala Lumpur, Malaysia
[c] Trauma Service, Department of Emergency Medicine, Kaohsiung Medical University Hospital, Taiwan

ARTICLE INFO

Keywords:
Delayed diagnosis
Emergency room (ER)
Affinity set
Data-mining
Topology
Rough set
Neural network
Support vector machine

ABSTRACT

At least 44,000 people die in hospitals each year as a result of medical errors, and these deaths are becoming the eighth-leading cause of death in the United States. Thus, medical providers have the responsibility to pay attention for reducing avoidable medical errors and improve patient safety as best as they can. It requires the rapid evaluation and prioritisation of life threatening injuries in the primary survey followed by a detailed secondary survey in the emergency room. However, time is always valuable and limited such that some important vital signs may be delayed and ignored. This research explores delayed diagnosis problem and uses the affinity set by Topology concept to classify/focus on key attributes causing delayed diagnosis (missed injury) in order to reduce error risk. Results interestingly indicate that when a patient can breathe normally, but his (or her) blood-pressure or pulse is abnormal, a high probability of delayed diagnosis exists. This affinity work also compares the performance with the model of rough set (Rosetta), neural network, support vector machine and logistic regression. And our affinity model shows its advantage by prediction accuracy and explanation power.

1. Introduction

Curing disease, maintaining health, and saving lives is the doctor's mission. People in traditional society believe the doctor's expertise is indisputable, trustworthy, and unmistakable. However, hospitals are not as safe as they should be, according to the results of two studies, the "Harvard Medical Practice Study" (Leape et al., 1991) and Institute of Medicine (IOM) investigation report "To Err Is Human" (Kohn, Corrigan, & Donaldson, 1999) published in 1986 and November 1999. The Harvard Study shows that 3.7% hospitalized patients suffered from medical injury, 27.6% suffered from medical negligence, 69% due to human error, 2.6% patients suffered permanent disability and 13.6% of medical errors led to death. On the other hand, the IOM report also shows that at least 44,000 people and perhaps as many as 99,000 people die in hospitals each year resulting from avoidable medical errors (Leape et al., 1991). These reports show that people dying in a given year due to avoidable medical errors exceed those of motor vehicle accidents (43,458), breast cancer (42,297), and AIDS (16,516) (Leape et al., 1991). Medical error deaths are becoming the eighth-leading cause of death in the United States (Kohn et al., 1999).

The aforementioned observations point out that reducing medical errors is critical when most are preventable (Department of Health). Diagnostic delay not only has clinical implications in terms of worsened outcome and potential long-term disability, but may also have financial and medico-legal consequences (Aaland & Smith, 1996; Brooksa, Holroydb, & Riley, 2004; Thomson & Greaves, 2008). Actually, real data of missed injury is hard to be defined and collected. Many researches indicate there is a strong link between the diagnosis delay and missed injury: when a patient is delayed, there is greater probability to generate medical errors (Born, Ross, & Iannacone, 1989; Furnival, Woodward, & Schunk, 1996). Therefore, this study explores diagnosis delay or failure of a planned action in operation in order to reduce the error risk. Diagnosis delay means patients' injuries are ignored or missed in emergency room (ER), but are identified by doctors in intensive care unit (ICU). The purpose of this paper is to find key attributes, which may lead to the delayed diagnosis problem by affinity set data-mining. The affinity set (Chen & Larbani, 2006; Larbani & Chen, 2008) is inspired from the vague interaction between people in social sciences (Freeman, 2004; Ho, 1998; Hwang, 1987; Luo, 2000), developed by Prof. Larbani and Prof. Chen as the data-mining tool to classify, analyze, and build the relationship between observed outcomes (consequences) and possible incomes (causes) of an information system. This research collects clinical data from ER in Chung-Ho Memorial Hospital, Kaohsiung Medical University, Taiwan; then uses the affinity data-mining model to identify key attributes leading to delayed diagnosis. Affinity mining results are also compared with some popular approaches for their

* Corresponding author. Tel.: +886 4 8511888x4120; fax: +886 4 8511270.
E-mail address: profchen@mail.dyu.edu.tw (Y.-W. Chen).

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.02.020

performances, and our affinity model has the advantage over the model of rough set (Rosetta) (Pawlak, 1982, 1991), neural network (Zbikowski & Hunt, 1996), support vector machine (Cortes & Vapnik, 1995) and logistics regression (Collett, 2003).

This paper is organized as follows: Section 2 introduces the basic concepts and definitions of affinity set; after that, the affinity data-mining model is proposed. Section 3 uses actual samples from Kaohsiung Medical University hospital, Taiwan to validate our affinity data-mining idea, deriving key delayed diagnosis attributes. This affinity work also competes with the model of rough set, neural network, support vector machine and logistic regression. Finally, Section 4 gives conclusions and recommendations based on our current achievements of the affinity data-mining model.

2. Technical background

Here, the basic concepts and definitions of affinity are simply reviewed (Larbani & Chen, 2008).

2.1. Affinity set

Affinity forms the basis for many aspects of social behavior, especially, the formation and evolution of social groups or social networks (Freeman, 2004; Ho, 1998; Hwang, 1987; Luo, 2000). Affinity has two meanings: the first is a natural liking for or attraction to a person, thing, idea, etc. Friendship and liking a kind of music or sport are examples of such affinity. This kind of affinity is called *direct affinity* in this paper. Direct affinity requires two elements to take place: the subjects between whom the affinity takes place and the affinity itself. The second defines affinity as a close relationship between people or things that have similar appearances, qualities, structures, properties or features. This paper calls it *indirect affinity*. For example, two people in the same political party have affinity between them; a group of people graduating from the same university also have affinity between them. Such affinity necessitates the presence of at least two elements: the subjects between whom there is affinity and the object (or medium) of the affinity. In this paper, we only apply the concept of *indirect affinity*.

So far as we know, in the literatures, there is no theory dealing with affinity as a vague and time-dependent concept, and little scholarly awareness that such a simple affinity idea could be developed for valuable models in information sciences. Fuzzy set theory could the best tool for representing vague and imprecise concepts so far; however, the affinity set proposed here is not merely a fuzzy set because assuming any type of membership function here is unnecessary. Instead, this work uses the closeness or distance between any two objects to develop useful models. Distance, is not a new idea; for example, the distance-based concept is used in ordinal preference ranking (Cook, 2006), in measuring similarity between fuzzy sets (Fonck, Fodor, & Roubens, 1998), and in fuzzy partition and fuzzy ordering applications (Ovchinnikov, 1991). The interesting and innovative idea in this study is using a decision maker's perception of distance/closeness to form his or her preferred affinity set. This new relation theory: affinity set theory is quite general, not only able to describe the similarity between objects, but also able to represent general relationships between objects, e.g., closeness, belongingness, equivalence, ..., etc., so that a decision maker can easily use this simple concept for modeling in information sciences.

The distance-based concept in the proposed affinity concept is certainly and strongly related to Topology science (Kelley, 1975; Mendelson, 1990); however, these topological abstractions are translated/simplified into useful modeling concepts and procedures here. Affinity vagueness and imprecision result from naturally vague and imprecise human judgment, evaluation, and feelings. The current framework makes a distinction between direct and indirect affinities. Direct affinity is expressed as a time-dependent fuzzy relation.

Definition 2.1. By affinity set we mean any object (real or abstract) that creates affinity between objects.

Some examples are given to clarify our idea.

Example 2.1. An institution or company is an affinity set, for it is an object that creates affinity between people that make them work together.

Definition 2.2. Let e and A be a subject and an affinity set, respectively. Let I be a subset of the time axis $[0, +\infty[$. The affinity between e and A is represented by a function

$$M_A^e(.) : I \to [0, 1]$$
$$t \to M_A^e(t).$$

The value $M_A^e(t)$ expresses the degree of affinity between the subject e and the affinity set A at time t. When $M_A^e(t) = 1$ this means that affinity degree of e with affinity set A is at the maximal level at time t; but $M_A^e(t) = 1$ does not mean that e belongs to A, unless the considered affinity measurement $M_A^e(t)$ is a function of belongingness degree. When $M_A^e(t) = 0$ this means that e has no affinity with A at time t. When $0 < M_A^e(t) < 1$, this means that e has partial affinity with A at time t. Here we emphasize the fact that the notion of affinity is more general than the notion of membership or belongingness: the later 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

$$M_U^e(.) : [0, +\infty[\to [0, 1]$$
$$t \to M_A^e(t)$$

and $M_U^e(t) = 1$, for all existing objects at time t, and for all times t.

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, some traditional referential set V, such as that when an object e is not in V, $M_A^e(t) = 0$ for all t in $I \subset [0,+\infty[$, 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 V. Thus, in the remainder of the paper when we refer to an affinity set, we assume that sets V and I are given.

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

$$F_A(.,.) : V \times I \to [0, 1] \quad (1)$$
$$(e, t) \to F_A(e, t) = M_A^e(t)$$

An element in real situations often belongs to a set for some time and not in that set other times. Such behavior can be represented using the affinity set notion. 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 V, the function (1) defining the affinity set A reduces to the fuzzy set describing degree variation of affinity of the element e over time.