Knowledge-Based Systems 74 (2015) 133–150

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Intuitionistic fuzzy recommender systems: An effective tool for medical diagnosis

Le Hoang Son *, Nguyen Tho Thong

VNU University of Science, Vietnam National University, Viet Nam

ARTICLE INFO

Article history:
Received 12 May 2014
Received in revised form 6 October 2014
Accepted 10 November 2014
Available online 20 November 2014

Keywords:
Accuracy
Fuzzy sets
Intuitionistic fuzzy collaborative filtering
Intuitionistic fuzzy recommender systems
Medical diagnosis

ABSTRACT

Medical diagnosis has been being considered as one of the important processes in clinical medicine that determines acquired diseases from some given symptoms. Enhancing the accuracy of diagnosis is the centralized focuses of researchers involving the uses of computerized techniques such as intuitionistic fuzzy sets (IFS) and recommender systems (RS). Based upon the observation that medical data are often imprecise, incomplete and vague so that using the standalone IFS and RS methods may not improve the accuracy of diagnosis, in this paper we consider the integration of IFS and RS into the proposed methodology and present a novel intuitionistic fuzzy recommender systems (IFRS) including: (i) new definitions of single-criterion and multi-criteria IFRS; (ii) new definitions of intuitionistic fuzzy matrix (IFM) and intuitionistic fuzzy composition matrix (IFCM); (iii) proposing intuitionistic fuzzy similarity matrix (IFSM), intuitionistic fuzzy similarity degree (IFSD) and the formulas to predict values on the basis of IFSD; (iv) a novel intuitionistic fuzzy collaborative filtering method so-called IFCF to predict the possible diseases. Experimental results reveal that IFCF obtains better accuracy than the standalone methods of IFS such as De et al., Szmidt and Kacprzyk, Samuel and Balamurugan and RS, e.g. Davis et al. and Hassan and Syed.

1. Introduction

In this section, we formulate the medical diagnosis problem and give some illustrated examples in Section 1.1. Section 1.2 describes the relevant works using the intuitionistic fuzzy sets for the medical diagnosis problem. Section 1.3 summarizes the limitations of those relevant works, and based on these facts the motivation and ideas of the proposed approach are highlighted in Section 1.4. Section 1.5 demonstrates our contributions in details, and their novelty and significance are discussed in Section 1.6. Lastly, Section 1.7 elaborates the organization of the paper.

1.1. The medical diagnosis problem

Medical diagnosis has been being considered as one of the most important and necessary processes in clinical medicine that determines acquired diseases of patients from given symptoms. According to Kononenko [20], diagnosis commonly relates to the probability or risk of an individual developing a particular state of health over a specific time, based on his or her clinical and non-clinical profile. It is useful to minimize the risk of associated health complications such as osteoporosis, small bowel cancer and increased risk of other autoimmune diseases. Mathematically, its definition is stated as follows.

Definition 1 (*Medical diagnosis*). Given three lists: $P = \{P_1, \ldots, P_n\}$, $S = \{S_1, \ldots, S_m\}$ and $D = \{D_1, \ldots, D_k\}$ where P is a list of patients, S a list of symptoms and D a list of diseases, respectively. Three values $n,\ m,\ k \in N^{+}$ are the numbers of patients, symptoms and diseases, respectively. The relation between the patients and the symptoms is characterized by the set- $R_{PS} = \{R^{PS}(P_i, S_j)|\forall i = 1, \ldots, n; \forall j = 1, \ldots, m\}$ where $R^{PS}(P_i, S_j)$ shows the level that patient P_i acquires symptom S_j and is represented by either a numeric value or a (intuitionistic) fuzzy value depending on the domain of the problem. Analogously, the relation between the symptoms and the diseases is expressed as $R_{SD} = \{R^{SD}(S_i, D_j)|\forall i = 1, \ldots, m; \forall j = 1, \ldots, k\}$ where $R^{SD}(S_i, D_j)$ reflects the possibility that symptom S_i would lead to disease D_j. The medical diagnosis problem aims to determine the relation between the patients and the diseases described by the set- $R_{PD} = \{R^{PD}(P_i, D_j)|\forall i = 1, \ldots, n; \forall j = 1, \ldots, k\}$ where $R^{PD}(P_i, D_j)$ is either 0 or 1 showing that patient P_i acquires disease D_j or not. The medical diagnosis problem can be shortly represented by the implication $\{R_{PS}, R_{SD}\} \rightarrow R_{PD}$.

* Corresponding author at: 334 Nguyen Trai, Thanh Xuan, Hanoi, Viet Nam. Tel.: +84 904171284; fax: +84 0438623938.
E-mail addresses: sonlh@vnu.edu.vn, chinhson2002@gmail.com (L.H. Son).

http://dx.doi.org/10.1016/j.knosys.2014.11.012
0950-7051/

Example 1. Consider the dataset in [31] having four patients namely P = {Ram, Mari, Sugu, Somu}, five symptoms S = {Temperature, Headache, Stomach-pain, Cough, Chest-pain} and five diseases D = {Viral-Fever, Malaria, Typhoid, Stomach, Heart}. The relations between the patients – the symptoms and the symptoms – the diseases are illustrated in Tables 1 and 2, respectively.

The relation between the patients and the diseases determined by the medical diagnosis is illustrated in Table 3. Since the domain of the problem is the intuitionistic fuzzy values, this relation is also expressed in this form. The most acquiring disease that the patients suffer is expressed in Table 4, which is converted from Table 3 by a trivial defuzzification method considering the maximal membership degree of disease among all.

Medical diagnosis is considered as an efficient support tool for clinicians to make the right therapeutical decisions especially in the cases that medicine extends its predictive capacities using genetic data [5]. As being observed in Table 3, medical diagnosis could assist the clinicians to enumerate the possible diseases of patients accompanied with certain membership values. Thus, it is convenient for clinicians, who are experts in this field, to quickly diagnose and give proper medicated figures. This fact clearly shows the importance of medical diagnosis in medicine sciences nowadays.

Table 1
The relation between the patients and the symptoms – R_{PS}.

P	Temperature	Headache	Stomach_pain	Cough	Chest_pain
Ram	(0.8, 0.1)	(0.6, 0.1)	(0.2, 0.8)	(0.6, 0.1)	(0.1, 0.6)
Mari	(0, 0.8)	(0.4, 0.4)	(0.6, 0.1)	(0.1, 0.7)	(0.1, 0.8)
Sugu	(0.8, 0.1)	(0.8, 0.1)	(0, 0.6)	(0.2, 0.7)	(0, 0.5)
Somu	(0.6, 0.1)	(0.5, 0.4)	(0.3, 0.4)	(0.7, 0.2)	(0.3, 0.4)

Table 2
The relation between the symptoms and the diseases – R_{SD}.

S	Viral_fever	Malaria	Typhoid	Stomach	Heart
Temperature	(0.4, 0)	(0.7, 0)	(0.3, 0.3)	(0.1, 0.7)	(0.1, 0.8)
Headache	(0.3, 0.5)	(0.2, 0.6)	(0.6, 0.1)	(0.2, 0.4)	(0, 0.8)
Stomach_pain	(0.1, 0.7)	(0, 0.9)	(0.2, 0.7)	(0.8, 0)	(0.2, 0.8)
Cough	(0.4, 0.3)	(0.7, 0)	(0.2, 0.6)	(0.2, 0.7)	(0.2, 0.8)
Chest_pain	(0.1, 0.7)	(0.1, 0.8)	(0.1, 0.9)	(0.2, 0.7)	(0.8, 0.1)

Table 3
The relation between the patients and the diseases – R_{PD} expressed by intuitionistic fuzzy values.

P	Viral_fever	Malaria	Typhoid	Stomach	Heart
Ram	(0.4, 0.1)	(0.7, 0.1)	(0.6, 0.1)	(0.2, 0.4)	(0.2, 0.6)
Mari	(0.3, 0.5)	(0.2, 0.6)	(0.4, 0.4)	(0.6, 0.1)	(0.1, 0.7)
Sugu	(0.4, 0.1)	(0.7, 0.1)	(0.6, 0.1)	(0.2, 0.4)	(0.2, 0.5)
Somu	(0.4, 0.1)	(0.7, 0.1)	(0.5, 0.3)	(0.3, 0.4)	(0.3, 0.4)

Table 4
The most acquiring diseases of patients.

P	Viral_fever	Malaria	Typhoid	Stomach	Heart
Ram	0	1	0	0	0
Mari	0	0	0	1	0
Sugu	0	1	0	0	0
Somu	0	1	0	0	0

1.2. The previous works

Computerized techniques for medical diagnosis such as fuzzy set, genetic algorithms, neural networks, statistical tools and recommender systems aiming to enhance the accuracy of diagnosis have been being introduced widely [20]. Nonetheless, an important issue in medical diagnosis is that the relations between the patients – the symptoms (R_{PS}) and the symptoms – the diseases (R_{SD}) are often vague, imprecise and uncertain. For instance, doctors could faced with patients who are likely to have personal problems and/or mental disorders so that the crucial patients' signs and symptoms are missing, incomplete and vague even though the supports of patients' medical histories and physical examination are provided within the diagnosis. Even if information of patients are clearly provided, how to give accurate evaluation to given symptoms/diseases is another challenge requiring well-trained, copious-experienced physicians. These evidences raise the need of using fuzzy set or its extension to model and assist the techniques that improve the accuracy of diagnosis. The definition of fuzzy set is stated below.

Definition 2. A *Fuzzy Set* (FS) [49] in a non-empty set X is a function

$$\begin{aligned} &\mu : X \to [0,1] \\ &x \mapsto \mu(x), \end{aligned} \quad (1)$$

where $\mu(x)$ is the membership degree of each element $x \in X$. A fuzzy set can be alternately defined as,

$$A = \{\langle x, \mu(x)\rangle | x \in X\}. \quad (2)$$

An extension of FS that is widely applied to the medical prognosis problem is *Intuitionistic Fuzzy Set* (IFS), which is defined as follows.

Definition 3. An *Intuitionistic Fuzzy Set* (IFS) [4] in a non-empty set X is,

$$\widetilde{A} = \left\{ \left\langle x, \mu_{\widetilde{A}}(x), \gamma_{\widetilde{A}}(x) \right\rangle | x \in X \right\}, \quad (3)$$

where $\mu_{\widetilde{A}}(x)$ and $\gamma_{\widetilde{A}}(x)$ are the membership and non-membership degrees of each element $x \in X$, respectively.

$$\mu_{\widetilde{A}}(x), \gamma_{\widetilde{A}}(x) \in [0,1], \quad \forall x \in X, \quad (4)$$

$$0 \leqslant \mu_{\widetilde{A}}(x) + \gamma_{\widetilde{A}}(x) \leqslant 1, \quad \forall x \in X. \quad (5)$$

The intuitionistic fuzzy index of an element showing the non-determinacy is denoted as,

$$\pi_{\widetilde{A}}(x) = 1 - \mu_{\widetilde{A}}(x) + \gamma_{\widetilde{A}}(x), \quad \forall x \in X. \quad (6)$$

When $\pi_{\widetilde{A}}(x) = 0$ for $\forall x \in X$, IFS returns to the FS set of Zadeh.

Some extensions of fuzzy sets are not appropriate for modeling uncertainty in the medical diagnosis such as the rough set [28], rough soft sets [11,12,16], intuitionistic fuzzy rough sets [50] and soft rough fuzzy sets & soft fuzzy rough sets [23]. The limitations of these sets, as pointed out by Yao [48], Rodriguez et al. [30], Jafarian and Rezvani [17] and many other authors lie to their intrinsic nature and how they are organized and operated such as (i) The positive and the boundary rules are considered in rough sets and their variants so that in cases of many concepts, the negative rules would be redundant; (ii) The modeling of linguistic information is limited due to the elicitation of single and simple terms that should encompass and express the information provided by the experts regarding the a linguistic variable; (iii) if exact membership degrees cannot be determined due to insufficient information then it is impossible to consider the uncertainty on the membership