Expert Systems With Applications 45 (2016) 97–107

Contents lists available at ScienceDirect

Expert Systems With Applications

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

A novel similarity/dissimilarity measure for intuitionistic fuzzy sets and its application in pattern recognition

Hoang Nguyen*

Department of Engineering Sciences, Gdynia Maritime University, Morska 83-87, 81–225 Gdynia, Poland

ARTICLE INFO

Keywords:
Intuitionistic fuzzy sets
Similarity measure
Dissimilarity
Knowledge
Fuzziness

ABSTRACT

Among the most interesting measures in intuitionistic fuzzy sets (IFSs) theory, the similarity measure is an essential tool to compare and determine degree of similarity between IFSs. Although there exist many similarity measures for IFSs, most of them cannot satisfy the axioms of similarity measure or provide reasonable results. In this paper, a novel knowledge-based similarity/dissimilarity measure between IFSs is proposed. Firstly, we define a new knowledge measure of information conveyed by the IFS and prove some properties of the proposed knowledge measure. Based on the proposed knowledge measure of IFSs, we construct a novel similarity/dissimilarity measure between IFSs and prove some properties of the proposed similarity measure. Then we use some illustrative examples to show that the proposed measures, though simple in concept and calculus, can overcome the drawbacks of the existing measures. Finally, we apply the proposed similarity/dissimilarity measure between IFSs in the pattern recognition problems to demonstrate that the proposed measure is the most reliable to deal with the pattern recognition problem in comparison with the existing similarity measures.

1. Introduction

In 1965, the theory of fuzzy set (FS) was first presented by Zadeh (1965) to deal with uncertain information. As a generalization of fuzzy set, intuitionistic fuzzy set (IFS) was introduced by Atanassov (1986, 1994) to deal with uncertainty of imperfect information. Since IFS presents information in terms of membership degree, non-membership degree and hesitancy degree, it is found to be more useful to deal with vagueness and uncertainty than that of FS. Many researchers have been trying to find a proper measure for IFSs, in order to evaluate and compare them. Among the most interesting measures in IFSs theory, similarity measure is an essential tool to compare and determine degree of similarity between two IFSs. Measuring similarity between IFSs has been intensively explored for decades and many similarity measures have been proposed in both theory and application aspects.

As a theory approach, Chen (1995) first proposed some similarity measures between vague sets. However, Hong and Kim (1999) pointed out some unreasonable cases of Chen's measures and proposed a set of modified measures. Later on, Li and Chen (2002) proposed some new similarity measures and their application in pattern recognition problems. Nevertheless, Liu in (Liu, 2005) showed that Li and Chen's methods have the same drawbacks as Chen's methods (Chen, 1995) and based on the distance measures proposed by Szmidt and Kacprzyk (2000), presented several new similarity measures between IFSs. The distance measure is often used in similarity measures between IFSs by researchers. Hung and Yang (2004) adopted the Hausdorff distance and developed several similarity measures for linguistic evaluations. Xu and Chen (2008) gave a comprehensive overview of the existing distance and similarity measures for IFSs and then proposed a series of distance and similarity measures based on the weighted Hamming distance, the weighted Euclidean distance and the weighted Hausdorff distance. Additionally, some studies on relationships between distance measure, similarity measure and entropy measure of IFSs have been made. Zeng and Guo (2008) investigated relationships of the normalized distance, the similarity measure, the inclusion measure and entropy of interval-valued fuzzy sets. They showed that the similarity measure, the inclusion measure and entropy of interval-valued fuzzy sets could be induced by the normalized distance of interval-valued fuzzy sets based on their axiomatic definitions. Besides, there have been some other types of similarity measure for IFSs. Boran and Akay (2014) proposed a new type of similarity measure for IFSs with two parameters, expressing L_p norm and the level of uncertainty, respectively. Subsequently, Intarapaiboon (2014) presented a set of theoretic similarity measures and their combination with a concept of lattice. Song, Zhu, and Chen (2014) proposed a novel probabilistic correlation-based similarity measure on text records with application in text matching. Montes, Pal, Janis, and Montes (2015) introduced an axiomatic definition of divergence

* Tel.: +48605565612.
E-mail address: van6302@wp.pl, hoang@am.gdynia.pl

http://dx.doi.org/10.1016/j.eswa.2015.09.045
0957-4174/

measures for IFSs, which are particular cases of dissimilarities between IFSs and investigated relationships among intuitionistic fuzzy divergences, dissimilarities and distances.

The similarity measures of IFSs are widely used in many applications such as medical diagnosis, decision making, pattern recognition and so on. While application to medical diagnosis, Ye (2011) conducted a study of the existing similarity measures between IFSs and proposed a cosine similarity measure, a weighted cosine similarity measure of IFSs. Similarly, Maoying (2013) developed a fuzzy cotangent similarity and weighted cotangent similarity measure between IFSs for medical diagnosis. Moreover, Davarzani and Khorheh (2013) proposed four new distance measures for IFSs and an application to the medical diagnosis progress in bacillus colonies recognition. As an application to the decision making problem, Xu (2007) introduced the concepts of positive-negative ideal IFS and extended some similarity measures to solve multi-attribute decision making problems. Subsequently, Xu and Yager (2009) proposed a new similarity measure between IFSs and applied it for consensus analysis in group decision making based on intuitionistic fuzzy preference relations. To solve the pattern recognition problems, Dengfeng and Chuntian (2002) proposed an axiomatic definition of similarity measure between IFSs based on high membership and low membership functions. However, Mitchell (2003) showed that the Dengfeng and Chuntian's similarity measure had some counterintuitive cases and modified the similarity measure based on statistical point of view. Next, Liang and Shi (2003) also considered some examples to show that the Dengfeng and Chuntian's similarity measure had some unreasonable cases and then proposed several new similarity measures for IFSs. Wang and Xin (2005) analyzed the relations between similarity measure and distance measure and applied the distance measure to pattern recognitions. Li, Olson, and Qin (2007) gave a comparative analysis of several existing similarity measures between IFSs and summarized their counterintuitive cases by examples in pattern recognition. Papakostas, Hatzimichailidis, and Kaburlasos (2013) investigated the main theoretical and computational properties of the existing distance and similarity measures for IFSs, as well as the relationships between them from a viewpoint of pattern recognition. In the other hand, Zhang and Yu (2013) presented a new similarity measure between IFSs based on their transformation into the symmetric triangular fuzzy numbers. They indicated that the proposed method contained more information with much less loss of information. Similarly, Chen and Chang (2015) also presented a similarity measure based on transformation techniques and applied the proposed measure to deal with pattern recognition problems. They claimed that the proposed similarity measure can outperform the existing similarity measures in solving the pattern recognition problems.

Besides the fuzzy similarity and intuitionistic fuzzy similarity measures, the probabilistic-based similarity (correlation) measures have been widely used in real world problems, for example in image clustering and partitioning (Kappes et al., 2015), in text matching (Song et al., 2014) or in time-series detection (Gao, Jiang, Chen, & Han, 2009). In (Song et al., 2014) a probabilistic correlation-based method was successfully adopted for unstructured text record similarity evaluation, where approximate string matching techniques for full text retrieval, i.e. edit distance and cosine similarity fail in cases of various word orders or incomplete information formats. Kappes et al. (2015) applied a probabilistic correlation in image clustering by using the perturbed maximum A-Posteriori (MAP) point estimates (Bayesian inference) to calculate globally consistent approximations to marginal distributions, making it enable to close open contour parts caused by imperfect local detection in image partitioning. (Gao et al., 2009) successfully adopted a probabilistic-based correlation method in the time-series detection of the large-scale distributed system, which can discover both the spatial (across system measurements) and temporal (across observation time) correlations.

Although there exist several similarity measures between IFSs, many unreasonable cases are made by the such measures as presented in (Li et al., 2007; Szmidt & Kacprzyk, 2013; Tan & Chen, 2014). Li et al. (2007) showed that there always are counterintuitive examples in pattern recognition among these existing similarity measures and pointed out a reason of this drawback as non-considering hesitancy degree in IFSs. Szmidt and Kacprzyk (2013) analyzed several geometric similarity measures between the IFSs and concluded that the symmetry of the complement elements in description of the IFS element is necessary to attain intuitively reliable results. Tan and Chen (2014) analyzed most of published researches on similarity measures comprehensively and proved that all existing similarity measures have counterintuitive cases. Inspired by this, we present in this paper a new similarity measure between IFSs, based on the new knowledge measure that makes it capable to evaluate differences between IFSs and provides reliable results. The performance evaluation of the proposed measure is shown in illustrative examples, assessing how much the measure is reasonable, and indicating the accuracy of the measure in comparison with others.

2. Basic concepts and a review of the existing similarity measures for IFSs

In 1986, Atanassov generalized the concept of fuzzy sets given by Zadeh (1965) by using membership and non-membership functions to deal with uncertainty of imperfect information. For any elements x of the finite universe of discourse X, an IFS A is defined by (Atanassov, 1986):

$$A = \{\langle x, \mu_A(x),\ \nu_A(x)\rangle | x \in X\}, \tag{1}$$

where the functions $\mu_A: X \to [0,1]$ and $\nu_A: X \to [0,1]$ denote a degree of membership and degree of non-membership of the element $x \in X$ to the set A, respectively, such that:

$$0 \le \mu_A(x) +\ \nu_A(x) \le 1,\ \forall x \in X. \tag{2}$$

To measure hesitancy degree of an element to an IFS, Atanassov introduced a third function given by:

$$\pi_A(x) = 1 - \mu_A(x) -\ \nu_A(x),\quad \forall x \in X,\ 0 \le \pi_A(x) \le 1, \tag{3}$$

which is also called the intuitionistic fuzzy index or the hesitation margin. If $\pi_A(x) = 0,\ \forall x \in X$, then $\mu_A(x) +\ \nu_A(x) = 1$ and the intuitionistic fuzzy set A is reduced to an ordinary fuzzy set.

The concept of a complement of an IFS A, denoted by A^c is defined as (Atanassov, 1986):

$$A^c = \{\langle x,\ \nu_A(x), \mu_A(x), \pi_A(x)\rangle | x \in X\}. \tag{4}$$

For any IFSs $A\, and\, B\, in\, X$, the following operations can be found in (Atanassov, 1986, 1999):

(D1) $A \subseteq B\ iff \mu_A(x) \le \mu_B(x), \nu_A(x) \ge \nu_B(x),\ \forall x \in X$;
(D2) $A = B\ iff\ A \subseteq B$ and $B \subseteq A$;
(D3) $A \cup B = \{\langle x, \max(\mu_A(x), \mu_B(x)), \min(\nu_A(x), \nu_B(x))\rangle | x \in X\}$;
(D4) $A \preceq B$ called A less fuzzy than B, i.e. for $\forall x \in X$,
if $\mu_B(x) \le \nu_B(x)\, then\, \mu_A(x) \le \mu_B(x)\, and\, \nu_A(x) \ge \nu_B(x)$;
if $\mu_B(x) \ge \nu_B(x)$ then $\mu_A(x) \ge \mu_B(x)\, and\, \nu_A(x) \le \nu_A(x)$.

Szmidt and Kacprzyk introduced intuitionistic fuzzy entropy measure, which is an extension of the De Luca and Termini's axioms for fuzzy sets. The axioms of entropy measure for IFSs were formulated as follows (Szmidt & Kacprzyk, 2001):

Let IFS(X) denotes the set of all IFSs in X, a map E: IFS(X)→[0,1] is called the intuitionistic fuzzy entropy, if it satisfies the following properties:

(D5) $E(A) = 0$ iff A is a crisp set;
(D6) $E(A) = 1\ iff\ \mu_A(x) =\ \nu_A(x),\ \forall x \in X$;