Applied Soft Computing 17 (2014) 79–89

Contents lists available at ScienceDirect

Applied Soft Computing

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

Distinguishability of interval type-2 fuzzy sets data by analyzing upper and lower membership functions

Lorenzo Livi[a,*], Hooman Tahayori[b], Alireza Sadeghian[b], Antonello Rizzi[a]

[a] Department of Information Engineering, Electronics, and Telecommunications, SAPIENZA University of Rome, Via Eudossiana 18, 00184 Rome, Italy
[b] Department of Computer Science, Ryerson University, 350 Victoria Street, Toronto, ON M5B 2K3, Canada

ARTICLE INFO

Article history:
Received 25 May 2013
Received in revised form 28 October 2013
Accepted 16 December 2013
Available online 6 January 2014

Keywords:
Interval type-2 fuzzy sets
Similarity and dissimilarity measures
Distinguishability of interval type-2 fuzzy sets
Unconventional pattern classification

ABSTRACT

In this paper, we deal with the problem of classification of interval type-2 fuzzy sets through evaluating their distinguishability. To this end, we exploit a general matching algorithm to compute their similarity measure. The algorithm is based on the aggregation of two core similarity measures applied independently on the upper and lower membership functions of the given pair of interval type-2 fuzzy sets that are to be compared. Based on the proposed matching procedure, we develop an experimental methodology for evaluating the distinguishability of collections of interval type-2 fuzzy sets. Experimental results on evaluating the proposed methodology are carried out in the context of classification by considering interval type-2 fuzzy sets as patterns of suitable classification problem instances. We show that considering only the upper and lower membership functions of interval type-2 fuzzy sets is sufficient to (i) accurately discriminate between them and (ii) judge and quantify their distinguishability.

1. Introduction

Research in data-driven inductive modeling systems resulted in the development of numerous classifiers that can automatically handle patterns defined as points in $\mathcal{I} = \mathbb{R}^n$ [52]. The main objective of classification problems is to accurately classify some labeled patterns of a given test set, based on some a priori knowledge of their (unknown) underlying *data generating process* that is available in the form of a training dataset. The *classification model* of a dataset is synthesized through a suitable learning (i.e., optimization) algorithm, guided by a proper *objective function*. There are many interesting practical pattern recognition problems that are intuitively defined on *structured* or *unconventional* patterns. Examples of such patterns are segmented images [3], audio/video signals [34], and metabolic networks [53]. In the literature, there are different formal representations to model such patterns; for instance labeled graphs, sequences of objects, and fuzzy sets [33,3,8,46,7,24,37,21,22,19,36,23,2]. When dealing with structured or unconventional data, usually the input space $\mathcal{I}$ is not directly interpretable as a common metric (or measure) space. Consequently, well-known classification systems, such as neuro-fuzzy networks and (adaptive) fuzzy inference systems [35,25,27,26,16], cannot be directly applied since they usually rely on Euclidean geometric properties of the input space $\mathcal{I}$ (such as *angles* and *distances*).

Recently, to address these issues, two mainstream approaches are widely used, the *kernel* [38] and the *dissimilarity* [33] based techniques. With the kernel-based approach, an appropriate positive semi-definite (PD) kernel function $k : \mathcal{I} \times \mathcal{I} \to \mathbb{R}$ is used to measure the similarity of the input data. Since PD kernels are Mercer's functions, the well-known Mercer's theorem applies, enabling the possibility of relying on the so-called *kernel trick* [38]. With the dissimilarity-based approach, a *dissimilarity* measure $d : \mathcal{I} \times \mathcal{I} \to \mathbb{R}^+$ is used to construct the *dissimilarity matrix*, whereby a new prototype-based formal representation space of the input data is derived. It is worth mentioning that both $d(\cdot, \cdot)$ and $k(\cdot, \cdot)$, which are generally termed as *matching algorithms*, are intimately related, playing a fundamental role when dealing with structured or unconventional patterns [21,33].

In this paper, we investigate the problem of classifying *interval type-2 fuzzy sets* (IT2FSs) by means of a general and light-weighted matching algorithm; the algorithm computes the similarity value between IT2FSs that, however, can be easily converted into a dissimilarity. Since the interpretation of an IT2FS as a pattern is not well-established in the technical literature, we refer to IT2FSs as unconventional patterns. The contribution of the paper is twofold. First, we show that, using different configurations of the proposed matching algorithm, enables the design of accurate and robust

* Corresponding author. Tel.: +39 0644585745.
E-mail addresses: llivi@scs.ryerson.ca, lorenz.livi@gmail.com (L. Livi), htahayor@scs.ryerson.ca (H. Tahayori), asadeghi@ryerson.ca (A. Sadeghian), antonello.rizzi@uniroma1.it (A. Rizzi).
URLs: https://sites.google.com/site/lorenzlivi/ (L. Livi), http://www.scs.ryerson.ca/ asadeghi/ (A. Sadeghian), http://infocom.uniroma1.it/ rizzi/index.html (A. Rizzi).

1568-4946/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.asoc.2013.12.020

classifiers for IT2FSs. Second, and most importantly, we show that the classification results obtained in this setting can be used to evaluate the *distinguishability* property of IT2FSs. Intuitively, two fuzzy sets are distinguishable if their support on the input domain is diversely distributed. Mencar and Fanelli [28] have provided different definitions of distinguishability in the type-1 fuzzy set (T1FS) case, which include also a dissimilarity-based approach: "two diversely distributed T1FSs will have a higher dissimilarity value in comparison with the two T1FSs having an analogous distribution in terms of support". We elaborate on the same dissimilarity-based interpretation to characterize distinguishable IT2FSs, relying on the achieved test set classification accuracy as a quantitative and formal indicator of distinguishability; the higher the test set classification accuracy results, the better the distinguishability of the analyzed IT2FSs. This fact enables judging over the outcomes of different procedures that are applied on the same data for generating a collection of IT2FSs. To date, few methods have been proposed for modeling words (concepts expressed in natural language) with IT2FSs [18,7,44,46], but judging over their outcome is subjective – to compare generated IT2FSs, the authors have mainly used the term "look more reasonable". In this paper, we quantitatively and objectively compare the IT2FSs generated by the four methods described in references [18,7,44,46].

The remainder of the paper is organized as follows. After a brief introduction to IT2FSs in Section 2, in Section 3 we introduce the context of similarity (and dissimilarity) measures, reviewing the state-of-the-art similarity measures for IT2FSs. Section 4 describes the proposed matching algorithm for IT2FSs, which is used as the core component in the classification. In Section 5, the performed experimental evaluations are discussed. Finally Section 6 concludes the paper, giving directions for future research.

2. Review of interval type-2 fuzzy sets

Type-2 fuzzy set (T2FS) is proposed as an extension to the type-1 fuzzy set [56]. T2FS enables handling additional *levels of uncertainty* by introducing the *fuzzy membership function*, which characterizes the membership value of an element as a T1FS [29,47]. Despite various efforts on making the cost of using T2FSs affordable, e.g., see [47,41,48,50,12,14,11], to compensate the computational complexities of T2FSs some variations are proposed; notably interval T2FS (IT2FS) [30] and shadowed fuzzy set (SFS) [43,45].

In IT2FS, the membership grade of an element is an interval that enables modeling the first degree of uncertainty. However, due to the uniform distribution sitting on top of the intervals, in an IT2FS there is no way to discriminate different choices of membership degrees. SFSs [43,45], however, provide a framework with more freedom degrees for handling uncertainties than IT2FSs with lower computational complexity comparing to general T2FSs. SFSs are generated through redistribution of the fuzziness associated with fuzzy grades of T2FSs in shadowed sets [51]. In this paper, however, we elaborate on IT2FSs.

An IT2FS $\tilde{\mathcal{A}}$ defined on the universe of discourse $\mathcal{X}$ is formally represented as

$$\tilde{\mathcal{A}} = \{(x, \mu_{\tilde{\mathcal{A}}}(x)) \ | \ x \in \mathcal{X}, \mu_{\tilde{\mathcal{A}}}(x) = \{(u, 1) \ | \ \underline{\mu}_{\tilde{\mathcal{A}}}(x) \leq u \leq \bar{\mu}_{\tilde{\mathcal{A}}}(x), [\underline{\mu}_{\tilde{\mathcal{A}}}(x), \bar{\mu}_{\tilde{\mathcal{A}}}(x)] \subseteq \mathcal{U} = [0, 1]\}\}. \quad (1)$$

In (1), x is called *primary variable*, and $[\underline{\mu}_{\tilde{\mathcal{A}}}(x), \bar{\mu}_{\tilde{\mathcal{A}}}(x)]$ denotes the *interval valued membership grade* of x in $\tilde{A}$. $\mathcal{X}$, as well as $\mathcal{U}$, can be a continuous or a finite set, defining in turn continuous or finite interval type-2 fuzzy sets.

An IT2FS is fully characterized by its so-called Footprint of Uncertainty (FOU), defined as:

$$\mathrm{FOU}_{\tilde{\mathcal{A}}} = \bigcup_{x\in\mathcal{X}} (x, [\underline{\mu}_{\tilde{\mathcal{A}}}(x), \bar{\mu}_{\tilde{\mathcal{A}}}(x)]). \quad (2)$$

Throughout the paper we use IT2FS and FOU interchangeably. FOU, as can be observed from Eq. (2), is a bounded region depicting the uncertainties associated with the membership grades of $\tilde{\mathcal{A}}$. FOU is completely identified by two T1FSs, namely *upper membership function* (UMF) and *lower membership function* (LMF), that are defined as:

$$\mathrm{UMF}_{\tilde{\mathcal{A}}} = \bar{\mathrm{FOU}}_{\tilde{\mathcal{A}}} = \{(x, \bar{\mu}_{\tilde{\mathcal{A}}}(x)) \ \ | x \in \mathcal{X}\}, \quad (3)$$

$$\mathrm{LMF}_{\tilde{\mathcal{A}}} = \underline{\mathrm{FOU}}_{\tilde{\mathcal{A}}} = \{(x, \underline{\mu}_{\tilde{\mathcal{A}}}(x)) \ \ | x \in \mathcal{X}\}. \quad (4)$$

More detailed discussions on the T2FSs theory, their operations, and related applications can be found in [30,47,41,48,50,49,42,13,10,15,58,39,6,5].

3. Similarity and dissimilarity measures

A dissimilarity measure on $\mathcal{I}$ is a bounded symmetric function $d : \mathcal{I} \times \mathcal{I} \to \mathbb{R}$, such that $\exists d_0 \in \mathbb{R}, -\infty < d_0 \leq d(x, y) < \infty, \forall x, y \in \mathcal{I}$ (where usually $d_0 = 0$), and $d(x, x) = d_0, \forall x \in \mathcal{I}$. If in addition $d(\cdot, \cdot)$ satisfies the triangular inequality and $d(x, y) = d_0 \leftrightarrow x = y$, then it is called a metric dissimilarity measure (metric in short). Analogously, we can define the concept of similarity measure, since in fact the two concepts are intimately related [3,33]. When $d(\cdot, \cdot)$ assumes values within the unit interval [0, 1], we refer to it as a *normal* dissimilarity measure. The same argument is true for the similarity measure $s(\cdot, \cdot)$. Although formal requisites of (dis)similarity algorithms are important during the analysis of the problem at hand, the design of effective-in-practice (dis)similarity measures for pattern recognition applications remains mostly an engineering, strongly problem-dependent, challenge [33,21,24,37,3]. Pattern recognition and data mining problems usually deal with complex data types, which points out the necessity of defining ad hoc (dis)similarity measures satisfying fewer constraints.

3.1. Brief review of similarity measures for IT2FSs

Let $\tilde{\mathcal{F}}(\mathcal{X})$ denotes the set of all IT2FSs on $\mathcal{X}$. A similarity measure on $\tilde{\mathcal{F}}(\mathcal{X})$ is a bounded function $s : \tilde{\mathcal{F}}(\mathcal{X}) \times \tilde{\mathcal{F}}(\mathcal{X}) \to \mathbb{R}^+$, which satisfies the following four axioms [4,54,57,55]:

1 (symmetry) $\forall \tilde{\mathcal{A}}, \tilde{\mathcal{B}} \in \tilde{\mathcal{F}}(\mathcal{X}), \quad s(\tilde{\mathcal{A}}, \tilde{\mathcal{B}}) = s(\tilde{\mathcal{B}}, \tilde{\mathcal{A}})$;
2 (vanishing) $s(\mathcal{D}, \mathcal{D}^c) = 0$, where $\mathcal{D}$ is a crisp set and $\mathcal{D}^c$ is its complement;
3 (maximization) $\forall \tilde{\mathcal{E}} \in \tilde{\mathcal{F}}(\mathcal{X}), \quad s(\tilde{\mathcal{E}}, \tilde{\mathcal{E}}) = \max_{\tilde{\mathcal{A}}, \tilde{\mathcal{B}} \in \tilde{\mathcal{F}}(\mathcal{X})} s(\tilde{\mathcal{A}}, \tilde{\mathcal{B}})$;
4 (monotonicity) $\forall \tilde{\mathcal{A}}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}} \in \tilde{\mathcal{F}}(\mathcal{X})$, if $\tilde{\mathcal{A}} \subseteq \tilde{\mathcal{B}} \subseteq \tilde{\mathcal{C}}$, then $s(\tilde{\mathcal{A}}, \tilde{\mathcal{B}}) \geq s(\tilde{\mathcal{A}}, \tilde{\mathcal{C}})$ and $s(\tilde{\mathcal{B}}, \tilde{\mathcal{C}}) \geq s(\tilde{\mathcal{A}}, \tilde{\mathcal{C}})$.

Wu and Mendel in [54] have proposed a similarity measure for IT2FSs, called *vector similarity measure*, which reads as:

$$\mathbf{s}_v(\tilde{\mathcal{A}}, \tilde{\mathcal{B}}) = [s_1(\tilde{\mathcal{A}}, \tilde{\mathcal{B}}), s_2(\tilde{\mathcal{A}}, \tilde{\mathcal{B}})]^T, \quad (5)$$

where $s_1(\tilde{\mathcal{A}}, \tilde{\mathcal{B}})$ measures the similarity between the shapes of the IT2FSs $\tilde{\mathcal{A}}$ and $\tilde{\mathcal{B}}$. $s_2(\tilde{\mathcal{A}}, \tilde{\mathcal{B}})$, however, calculates the proximity of $\tilde{\mathcal{A}}$ and $\tilde{\mathcal{B}}$. Both $s_1(\cdot, \cdot)$ and $s_2(\cdot, \cdot)$ are based on the analysis of the centroid of the two input IT2FSs. The final scalar similarity value is then obtained by multiplying the outcomes of $s_1(\cdot, \cdot)$ and $s_2(\cdot, \cdot)$.