Knowledge-Based Systems 55 (2014) 114–124

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Encoding words into Cloud models from interval-valued data via fuzzy statistics and membership function fitting

Xiaojun Yang *, Liaoliao Yan, Hui Peng, Xiangdong Gao

Luoyang Electronic Equipment Test Center, Luoyang, He'nan 471003, China

ARTICLE INFO

Article history:
Received 28 March 2013
Received in revised form 9 October 2013
Accepted 12 October 2013
Available online 19 October 2013

Keywords:
Computing with words
Cloud model
Enhanced interval approach
Fuzzy statistics
Membership function fitting
Histogram

ABSTRACT

When constructing the model of a word by collecting interval-valued data from a group of individuals, both interpersonal and intrapersonal uncertainties coexist. Similar to the interval type-2 fuzzy set (IT2 FS) used in the enhanced interval approach (EIA), the Cloud model characterized by only three parameters can manage both uncertainties. Thus, based on the Cloud model, this paper proposes a new representation model for a word from interval-valued data. In our proposed method, firstly, the collected data intervals are preprocessed to remove the bad ones. Secondly, the fuzzy statistical method is used to compute the histogram of the surviving intervals. Then, the generated histogram is fitted by a Gaussian curve function. Finally, the fitted results are mapped into the parameters of a Cloud model to obtain the parametric model for a word. Compared with eight or nine parameters needed by an IT2 FS, only three parameters are needed to represent a Cloud model. Therefore, we develop a much more parsimonious parametric model for a word based on the Cloud model. Generally a simpler representation model with less parameters usually means less computations and memory requirements in applications. Moreover, the comparison experiments with the recent EIA show that, our proposed method can not only obtain much thinner footprints of uncertainty (FOUs) but also capture sufficient uncertainties of words.

1. Introduction

Human knowledge and perceptions are usually described in a natural language. Therefore, in 1996, Zadeh proposed the phrase "*computing with words*" (CWW) in his seminal paper [1]. CWW is "a methodology in which the objects of computation are words and propositions drawn from a natural language." Inspired by the remarkable human capability of using words, CWW will play a pivotal role in everyday's automated reasoning and decision making. Since Zadeh's seminal paper, many articles have been published with the phrase "computing with words" in their titles, such as [2–7].

Constructing the models of words is the first step in the CWW paradigm. Mendel pointed out that; words mean different things to different people, and so do uncertainties [2]. We, therefore, need a fuzzy set model for a word that has the potential to capture its uncertainties. In previous publications, words were modeled by type-1 fuzzy sets (T1 FSs) [1,4,8,9]; IT2 FSs [2,10,11] or general type-2 fuzzy sets (GT2 FSs) [12].

However; as claimed in [2,10–12], when surveying from a group of individuals, both interpersonal and intrapersonal uncertainties coexist. Unfortunately, a T1 FS does not have enough degrees of freedom to model both uncertainties about collecting word data from a group of individuals. So, recently, Mendel first proposed two approaches to encoding words into IT2 FSs from person membership function (MF) survey and interval endpoints survey respectively [2]. After that, two enhanced versions of the interval endpoints survey approach were presented in [10,11]. An assumption made by these methods is that each individual's data interval is uniformly distributed. First, the interval endpoints data are preprocessed in the data part to remove bad values. Then, make the mean and standard deviation of a T1 FS (triangle, left-, or right-shoulder) equate to those of a uniformly distributed data interval to obtain each individual's MF. At last, the FOU for the word is built from these embedded T1 FSs. These interval endpoints survey approaches to modeling words using IT2 FSs have shown great promise. More recently, to obtain a word model with no specification of a MF type and without loss of information, Miller et al. proposed an approach to translating interval-valued data into GT2 FSs [12]. Since the number of a GT2 FS's parameters is proportional to the number of surveyed individuals, the word model using a GT2 FS is not a parsimonious parametric model. It can result in huge numbers of computations and enormous memory requirements in applications. Instead, we should develop a more parsimonious parametric model for a word. A parsimonious parametric model with less parameters usually means faster convergence in the calculation. Parsimony is achieved by choosing a model with the minimum number of parameters that best approximates the data [2].

* Corresponding author. Tel.: +86 13613799712.
E-mail address: yangxiaojun2007@gmail.com (X. Yang).

0950-7051/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2013.10.014

In this paper, we propose a novel and simple approach to modeling words with normal Cloud models, referred to as encoding words into Cloud models using fuzzy statistics and MF fitting. The Cloud model [13–15] defined by Li et al. is a cognitive model that mainly reflects the uncertainties of things in the universe and the concepts in human knowledge. The uncertainties include the fuzziness, randomness, and the association between them. Several kinds of Cloud models were defined, including normal Cloud model, trapezoidal Cloud model, triangular Cloud model, half Cloud model, and combined Cloud model. Thus far, the normal Cloud model based on normal distribution and Gaussian MF is the most applicable one. It has been successfully applied to many fields, such as data mining [13,16], uncertainty reasoning [17,18]; time series prediction [19,20]; image processing [21,22]; representation of human knowledge and subjective judgments in Delphi method and analytic hierarchy process [23–25]. In particular, the normal Cloud model is characterized by only three parameters: Expectation (*Ex*), Entropy (*En*), and Hyper-entropy (*He*). We therefore develop a much more parsimonious parametric model for a word based on the normal Cloud model. Moreover, experimental results show that, we can not only obtain much thinner FOUs but also capture sufficient uncertainties of words.

The rest of the paper is organized as follows. Section 2, as the background, introduces the concept and characteristics of the Cloud model and reviews the EIA [11]. Section 3 describes the framework of our proposed method. Section 4 presents some examples to demonstrate the performance of our proposed method, and compares and discusses the experimental results. Section 5 concludes this paper.

2. Background

In this section we provide some brief background information on the techniques employed in this article, in particular the normal Cloud model [14] and the EIA [11].

2.1. Cloud model

The aims of (I)T2 FSs and Cloud models are both to deal with the uncertainty of the MF which traditional T1 FSs do not consider. According to the T1 FS theory, once the MF is determined, one and only one accurate membership degree can be calculated for any given element in the universe to measure the uncertainty of this element belonging to the associated concept. This is obviously inconsistent with the spirit of the fuzzy set because the uncertainty of an element belonging to a fuzzy concept becomes certain and precise in this case. Thus, Li et al. defined the Cloud model by considering whether allowing a stochastic disturbance of the membership degree encircling a determined central value is more feasible. Among several kinds of defined Cloud models, the normal Cloud model based on normal distribution and Gaussian MF is the most commonly used one. It can express many uncertain phenomena in both natural and social sciences. In this paper, we discuss only the normal Cloud model, and the name of Cloud model is regarded as equivalent to the normal Cloud model. The normal Cloud model is defined as follows.

Definition 1 (*[14,15]*). Let U be the universe of discourse and C be a qualitative concept in U. If $x \in U$ is a random instantiation of concept C, x satisfies $x \sim N(Ex, En'^2)$, where $En' \sim N(En, He^2)$, and the certainty degree of x belonging to concept C is

$$\mu = e^{-\frac{(x-Ex)^2}{2(En')^2}}, \tag{1}$$

then the distribution of x in the universe U is called a *normal Cloud*, and every x with the certainty degree μ is defined as a Cloud drop.

The Cloud model can effectively integrate the randomness with fuzziness of cognitive concepts and can describe the overall quantitative property of a concept by three numerical characteristics, namely, *Ex*, *En*, and *He*. *Ex* is the mathematical expectation of the Cloud drops belonging to a concept in the universe. It is the most representative and typical sample of the qualitative concept. *En* represents the uncertainty measurement of a qualitative concept. It is determined by both the randomness and the fuzziness of the concept. On the one hand, as the measurement of randomness, *En* reflects the dispersing extent of the Cloud drops, which is similar to the standard deviation of a random variable. On the other hand, as the measurement of fuzziness, it represents the scope of the universe that can be accepted by the concept. *He* is the uncertainty degree of *En*, also seen as the entropy of entropy. *He* reflects the dispersion of the Cloud drops. The larger the *He* is, the larger its dispersion is, the larger the randomness of degree of membership is, and the larger the thickness of the Cloud is.

The *Forward Normal Cloud Generator* is used to transform the Cloud model from its qualitative representation into its quantitative representation. It generates Cloud drops based on the Cloud parameters (*Ex*, *En*, *He*). The generating algorithm of the one-dimension normal Cloud *CG* (*Ex*, *En*, *He*, *n*) can be described as follows [15,25].

Algorithm 1. Forward Normal Cloud Generator $CG(Ex, En, He, n)$.

Input: Three parameters, *Ex*, *En*, and *He*, and the number of Cloud drops n

Output: n Cloud drops x and their certainty degrees μ, i.e., drop (x_i, μ_i), $i = 1, 2, \ldots, n$.

Step 1: A normal random number En'_i with expectation of *En* and a standard deviation of *He* is generated.
$En'_i = NORMRND\ (En, He)$.

Step 2: A normal random number x_i, with expectation of *Ex* and a standard deviation of En'_i is generated.
$x_i = NORMRND\ (Ex, En'_i)$.

Step 3: The certainty degree of x_i is calculated as follows:

$$\mu_i(x_i) = e^{-\frac{(x_i-Ex)^2}{2(En'_i)^2}}. \tag{2}$$

Step 4: x_i is a Cloud drop with the certainty degree μ_i expressed as drop (x_i, μ_i).

Step 5: Steps 1 to 4 are repeated until the Cloud drops generated are sufficient for n.

For example, Fig. 1 presents the result of *CG* (5.12, 1.06, 0.05, 5000). It can represent the qualitative concept "moderate amount" on the scale of 0–10.Let parameter *He* be equal to 0, then the Cloud model becomes a T1 FS with a Gaussian MF. We define this Gaussian MF as the expectation MF μ_{expec} of the Cloud model

$$\mu_{expec} = e^{-\frac{(x-Ex)^2}{2En^2}}. \tag{3}$$

As aforementioned, the Cloud model is defined by allowing a stochastic disturbance of the membership degree encircling the expectation MF, i.e., the membership degrees of each x in the universe of discourse may be multiple uncertain values encircling the expectation MF instead of a fixed value. According to the "3*He* rule", which is similar to the "3σ rule" of normal distribution, 99.7% of Cloud drops are contained between the upper MF μ_{upper}