Available online at www.sciencedirect.com
ScienceDirect
Information Sciences 177 (2007) 988–1006
INFORMATION SCIENCES
AN INTERNATIONAL JOURNAL
www.elsevier.com/locate/ins

Computing with words and its relationships with fuzzistics

Jerry M. Mendel *

Signal and Image Processing Institute, Department of Electrical Engineering, University of Southern California, 3740 McClintock Avenue, Los Angeles, CA 90089-2564, United States

Received 22 November 2005; received in revised form 28 May 2006; accepted 10 June 2006

Abstract

Words mean different things to different people, and so are uncertain. We, therefore, need a fuzzy set model for a word that has the potential to capture their uncertainties. In this paper I propose that an interval type-2 fuzzy set (IT2 FS) be used as a FS model of a word, because it is characterized by its footprint of uncertainty (FOU), and therefore has the potential to capture word uncertainties. Two approaches are presented for collecting data about a word from a group of subjects and then mapping that data into a FOU for that word. The *person MF approach*, in which each person provides their FOU for a word, is limited to fuzzy set experts because it requires the subject to be knowledgeable about fuzzy sets. The *interval end-points approach*, in which each person provides the end-points for an interval that they associate with a word on a prescribed scale is not limited to fuzzy set experts. Both approaches map data collected from subjects into a parsimonious parametric model of a FOU, and illustrate the combining of fuzzy sets and statistics—type-2 fuzzistics.

Keywords: Computing with words; Fuzzistics; Interval type-2 fuzzy sets; Type-2 fuzzy sets; Inverse problems

1. Introduction

In 1996 Zadeh [39] published an important paper that for the first time equated fuzzy logic (FL) with *computing with words* (CWW). He did not mean that computers would actually compute using words rather than numbers. He meant that fuzzy logic (e.g. [37,38]) would provide a natural framework for humans to interact with computers using words, and that the computer would provide words back to the humans. Since the publication of his paper, there have been many articles and even books bearing the phrase "computing with words". A small sampling of these are [11,12,14,16,17,35,40,41]. A related reference is [33]. FL is viewed in these publications as the machinery that will let "input" words, that are provided by a human, be transformed within the computer to "output" words, that are provided back to that human. This machinery may take the form of IF–THEN rules, a fuzzy weighted average, a fuzzy Choquet integral, etc., for which the established mathematics of fuzzy sets provides the transformation from the input words to the output words. Potential applications for CWW are many and include web-based searches, summarizations, subjective judgments, subjective decisions, etc.

* Tel.: +1 213 740 4445; fax: +1 213 740 4456.
E-mail address: Mendel@sipi.usc.edu

0020-0255/$ - see front matter
doi:10.1016/j.ins.2006.06.008

Words mean different things to different people, and so are uncertain. We, therefore, need a fuzzy set model for a word that has the potential to capture their uncertainties [9,18]. In this paper I propose that an interval type-2 fuzzy set (IT2 FS) be used as a FS model of a word, because it is characterized by its footprint of uncertainty (FOU), and therefore has the potential to capture word uncertainties. Why I propose to use an IT2 FS rather than a more general T2 FS will be discussed in Section 2.

Word FS models must be based on data that are collected from a group of subjects. How to do this, and how to then map the data into the parameters of a T1 MF has been reported on by a number of authors (e.g. [10]). Names for the different T1 methods include: polling [7,11], direct rating [1,26,27,31], reverse rating [26,29–32], interval estimation [3–5,42], and transition interval estimation [4]. Unfortunately, none of these methods transfers the uncertainties about collecting word data from a group of subjects into the MF of a T1 FS, perhaps because a T1 FS does not have enough degrees of freedom to do this; hence, I will not elaborate upon them in this paper.

In the rest of this paper I will describe two very different methods for collecting data about words from a group of people and then obtaining the FOU of an IT2 FS model for the word. Both methods require that:

1. A continuous scale is established for each variable of interest (sometimes a natural scale exists, e.g., as in pressure, temperature, volume, etc., otherwise, a natural scale does not exist, e.g., as in touching, eye-contact, beautiful, etc.).
2. A vocabulary of words is created that covers the entire scale.
3. Data are collected from a group of subjects about the words.

We will describe our methodologies for the continuous scale numbered 0–10. Of course, other scales could be used. An interesting research issue is whether or not data collected on one scale can be automatically re-scaled to another scale. Another very interesting research issue is whether or not data collected on a scale for one specific application can be re-scaled on that same or a different scale for another application. These issues are beyond the scope of this paper.

Every application has its vocabulary. A vocabulary of words should therefore be created for a specific application, and it should be large enough so that a human will feel linguistically comfortable in interacting with a computer. The vocabulary should have words that feel like they will range from 0 to 10, so that the entire scale will be covered by the vocabulary. Redundant words and coverage is not the issue here, although it could be an important issue when designing, e.g., an IF–THEN interface that maps the user's input words (which can come from the large vocabulary) into a computer's output word (which can also come from the large vocabulary). In practice, rules may only be created for the smallest sub-set of words that cover the entire scale, thereby keeping the number of rules as small as possible.

The first data collection method (Section 2) requires each person in the group of subjects to provide a *person MF* for a word (and for all words in the vocabulary). It is limited to people who are already knowledgeable about a fuzzy set and a MF—people we shall call *FS experts*. The second data collection method (Section 3) requires each person in the group of subjects to provide interval end-points for a word, on a scale of, e.g., 0–10. This method can be applied to anyone because it does not require a subject know anything about a FS or MF.

The term *fuzzistics*, which is a merging of the words *fuzzy* and *statistics*, was coined in [18] to summarize the problem of going from word data collected from a group of subjects, with its inherent random uncertainties that are quantified using statistics, to a word fuzzy set model that captures a measure of the word data-uncertainties. When the FS model is an IT2 FS we shall call this *type-2 fuzzistics*. Appendix A provides a brief introduction to IT2 FSs.

2. Person MF approach

2.1. Overview

In the person MF approach [18] we: (1) collect person MF data that reflects both the intra- and inter-levels of uncertainties about a word, from a group of people; (2) define the IT2 FS model for a word as a specific