Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Fuzzy Sets and Systems 157 (2006) 1662–1673
FUZZY sets and systems
www.elsevier.com/locate/fss

# Linguistic modelling based on semantic similarity relation among linguistic labels

Yongchuan Tang[a,*], Jiacheng Zheng[b]

[a] College of Computer Science, Zhejiang University, Hangzhou 310027, PR China
[b] College of Economics, Zhejiang University, Hangzhou 310027, PR China

Received 5 July 2005; received in revised form 23 February 2006; accepted 23 February 2006
Available online 31 March 2006

**Abstract**

In classical linguistic modelling for imprecise concepts and uncertain reasoning, a basic idea is to use fuzzy sets to describe the semantics of imprecise concepts, the theoretic foundation of this modelling is fuzzy logic. But in some practical applications, the determination of membership functions associated with vague concepts is difficult or impossible. In this paper, we present a new linguistic modelling technique based on the semantic similarity relation among linguistic labels. In this new linguistic model, a fuzzy relation is utilized to represent the semantic relation among linguistic labels. The elements in the fuzzy relation are interpreted as the degrees of similarities or the degrees of semantic overlapping between the corresponding linguistic labels. We develop an inference method for computing the degree of similarity between any two linguistic expressions which are the logic formulas generated by applying the logic connectives to the linguistic labels. This inference method utilizes the consonant mass assignment functions induced from the fuzzy similarity relation on linguistic labels. We show that the linguistic model presented in this paper satisfies some perfect properties, e.g. the excluded middle law, law of non-contradiction. This new linguistic modelling is illustrated finally in a linguistic decision-making example.


*Keywords:* Linguistic modelling; Linguistic label; Semantic similarity; Consonant mass assignment function; Fuzzy relation; Computing with words

## 1. Introduction

In modelling for linguistic uncertainty, linguistic labels are the appropriate tools to describe vague concepts in natural language. In natural language, the objects are often represented by the set of linguistic labels when uncertain reasoning is involved. One main objective of linguistic modelling or computing with words [23] in natural language is to combine all linguistic labels or descriptions on the underlying objects, and to make decision according to the linguistic information. Such linguistic modelling is of central importance for many of emerging information technology. For example, the linguistic modelling has important application in linguistic information fusion [4,7–9].

* Corresponding author. Tel./fax: +86 571 87952639.
*E-mail address:* yongchuan@263.net (Y. Tang).

0165-0114/$ - see front matter 
doi:10.1016/j.fss.2006.02.014

In classical linguistic modelling [2,3,15,16], the fuzzy set or membership function associated with each linguistic label is used to represent its semantic. Zadeh firstly introduced the concept of linguistic variable as a model of how words or labels can represent vague concepts in natural language [20–22]. The use of fuzzy sets is central to computing with words as they provide a means of modelling vagueness underlying most natural language terms [12,23]. But, in some practical applications, the determination of membership functions or fuzzy sets associated with linguistic labels is difficult or impossible. Some linguistic labels, such as grand, steady, are often applied to product design evaluation [10]. In fact, the linguistic uncertainty can be directly modelled by a fuzzy relation on the set of linguistic labels. This fuzzy relation can be interpreted as the degrees of semantic similarities among linguistic labels. This paper uses a fuzzy relation to describe the degrees of similarities among linguistic labels, instead of fuzzy sets to describe the semantics of linguistic labels. Of course, the fuzzy set can be considered as a relational description between the linguistic label and elements in the corresponding universe of discourse. It is also possible to build a fuzzy relation among linguistic labels based on the fuzzy sets associated with linguistic labels. This topic is not our focus of this paper. The main contribution of this paper is to directly build the linguistic model based on fuzzy relation among linguistic labels, but not fuzzy sets associated with linguistic labels.

The fuzzy relation among linguistic labels measures the degree of similarity between any two linguistic labels. Now, our question is how to infer the degree of similarity between any two linguistic expressions, here the linguistic expressions are linguistic formulas generated by applying the logical connectives $\neg$, $\wedge$, $\vee$ and $\rightarrow$ to the linguistic labels [13]. One solution for this question, an inference method based on mass assignment theory, is proposed in this paper. Firstly, for each linguistic label, a fuzzy set defined on the set of all linguistic labels is induced from the initial fuzzy relation, this fuzzy set is then converted to a consonant mass assignment function on the set of linguistic labels. By using the consonant mass assignment theory for a fuzzy set [1], the degree of similarity between any two linguistic expressions is inferred. Especially, we show that this linguistic model is consistent with the properties of two-valued logic. For example, the excluded middle law and non-contradiction law hold in this linguistic model.

This paper is organized as follows. Section 2 presents a model for linguistic uncertainty using a fuzzy relation among linguistic labels, and proposes an inference method for computing the degree of similarity between any two linguistic expressions. Section 3 gives some properties of proposed inference method, which can be used to simplify the inference process, and we show that this method is consistent with two-valued logic. Section 4 illustrates a linguistic information fusion example. Section 5 gives some discussions on this linguistic model. The last section is our conclusions.

## 2. Linguistic model based on fuzzy relation

In this paper, we concentrate on a finite set of linguistic labels $LA = \{L_1, \ldots, L_n\}$. A fuzzy relation $R = (r_{ij})_{n\times n}$ is defined on $LA$, where the element $r_{ij} \in [0, 1]$ in the $i$th row and $j$th column of $R$ represents the degree of semantic overlapping or similarity between linguistic labels $L_i$ and $L_j$, or $1 - r_{ij}$ represents the semantic difference between linguistic labels $L_i$ and $L_j$. It is reasonable to assume that the fuzzy relation $R$ satisfies $r_{ij} = r_{ji}$, and $r_{ii} = 1$ for $i, j = 1, \ldots, n$. In the following, 2-tuple $\langle LA, R\rangle$ is called as a linguistic model.

In the original definition on linguistic labels [20–22], a linguistic variable can attain its value in $LA$, and a semantic rule can associate a normalized fuzzy set with each linguistic label in $LA$. In addition, Zadeh [20–22] claimed that a new linguistic label can be formed by applying hedges to existing linguistic labels, and there is a simple functional relationship between the fuzzy sets associated with the linguistic label and new linguistic label generating from it. In our view, the semantic relation among linguistic labels is central to linguistic modelling in some situations. Since in the linguistic model $\langle LA, R\rangle$ there is no semantic device for the single linguistic label, we consider only fixed finite labels set where all the labels and their semantic relations are predefined.

Note that the linguistic label $L_i \in LA$ in our linguistic model can be still considered as a description on a vague concept, but no fuzzy set or membership function is defined for $L_i$. In some situations, the semantic relations among linguistic labels are enough to capture the vagueness in most natural language. For instance, if we have $LA = \{\textit{very good}, \textit{good}, \textit{moderate}, \textit{low}\}$, then on the view point of linguistic semantic relation, we think that *very good* is much closer to *good* than *moderate*, and *good* is much closer to *moderate* than *low*. Therefore, one possible definition of fuzzy relation $R$ on $LA$ satisfies $r(\textit{very good}, \textit{good}) = 0.75$, $r(\textit{very good}, \textit{moderate}) = 0.5$, $r(\textit{very good}, \textit{low}) = 0.25$, $r(\textit{good}, \textit{moderate}) = 0.75$, $r(\textit{good}, \textit{low}) = 0.5$, $r(\textit{moderate}, \textit{low}) = .75$.