Knowledge-Based Systems 88 (2015) 154–164

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Building and managing fuzzy ontologies with heterogeneous linguistic information

J.A. Morente-Molinera [a,*], I.J. Pérez [b], M.R. Ureña [a], E. Herrera-Viedma [a,c]

[a] Dept. of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain
[b] Dept. of Computer Sciences and Engineering, University of Cádiz, Cádiz, Spain
[c] Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia

A R T I C L E I N F O

Article history:
Received 4 May 2015
Received in revised form 22 June 2015
Accepted 23 July 2015
Available online 30 July 2015

Keywords:
Fuzzy linguistic modeling
Multi-granular linguistic information
Computing with words
Fuzzy ontology

A B S T R A C T

Fuzzy ontologies allow the modeling of real world environments using fuzzy sets mathematical environment and linguistic modeling. Therefore, fuzzy ontologies become really useful when the information that is worked with is imprecise. This happens a lot in real world environments because humans are more used to think using imprecise nature words instead of numbers. Furthermore, there is a high amount of concepts that, because of their own nature, cannot be measured numerically. Moreover, due to the fact that linguistic information is extracted from different sources and is represented using different linguistic term sets, to deal with it can be problematic. In this paper, three different novel approaches that can help us to build and manage fuzzy ontologies using heterogeneous linguistic information are proposed. Advantages and drawbacks of all of the new proposed approaches are exposed. Thanks to the use of multi-granular fuzzy linguistic methods, information can be expressed using different linguistic term sets. Multi-granular fuzzy linguistic methods can also allow users to choose the linguistic term sets that they prefer to formulate their queries. In such a way, user-computer communication is improved since users feel more comfortable when using the system.

1. Introduction

Ontologies have become an important tool in the domain modeling field. Thanks to them, it is possible to carry out real world representations, stablish axioms and obtain conclusions of them [1,2]. Ontologies have been wide used in several fields. In biomedicine field [3–5], ontologies have been employed, for example, to build knowledge databases about genes and proteins characteristics that help researchers to classify and understand how the human body works. In semantic web field [6–8], ontologies have been used to classify concepts that can be referred through the web. This way, searches are improved and give better results to the users because a concept, instead of a word that can have different meanings, is used. In the artificial intelligence field [9–11], ontologies can also be applied to create knowledge databases to be used in systems that employ the provided information to carry out different tasks.

However, classical Crisp Ontologies have one important drawback, that is, their element descriptions can only be expressed using crisp membership values. Consequently, each described element has a set of fulfilled characteristics and another one with characteristics that do not describe the element. That is, membership value of each element to each concept is represented by the values {0,1} where 0 means that the element does not fulfill the concept and 1 means that the element has the characteristic expressed by the concept. In real world problems, this kind of scenario is not enough to describe correctly certain situations. For solving this issue and being able to provide a more flexible way of carrying out descriptions, Fuzzy Ontologies (FO) has been developed. Thanks to FOs, it is possible to provide membership values from the defined elements to the concepts using the interval [0,1]. Therefore, each described element can fulfill concepts totally (1 value), do not fulfill it (0 value) or partially fulfill it with a certain degree value (]0,1[). Thanks to this new representation, it is possible to model the uncertainty that is implicit in many real world environments and. Using fuzzy sets theory [12], it is possible for the ontology to deal with it using its associated mathematical environment. FOs is a field that is clearly present in the recent literature as it can be seen in [13–15].

* Corresponding author.
E-mail addresses: jamoren@decsai.ugr.es (J.A. Morente-Molinera), ignaciojavier.perez@uca.es (I.J. Pérez), raquel@decsai.ugr.es (M.R. Ureña), viedma@decsai.ugr.es (E. Herrera-Viedma).

http://dx.doi.org/10.1016/j.knosys.2015.07.035
0950-7051/

FOs also open the way for introducing linguistic modeling in this research field [16]. Thanks to it, elements can be described by using words instead of numbers. Linguistic modeling and linguistic term sets (LTSs) [17] in order to describe elements have one main advantage and one main drawback. The advantage is that words are more flexible that numbers. Consequently, this is the best way when trying to model concepts whose meaning is imprecise. They are also easier for humans to use that numbers making them a perfect choice when trying to model people opinions. On the other hand, the main drawback of using linguistic labels is the loss of precision that they produce when trying to represent precise data values.

FOs are used to create big knowledge stores whose data can come from different information sources, and therefore, source information is expressed using different representation methods. Due to the heterogeneity of the information, sometimes it is difficult to manage it. In such a way, it is extremely important to be able to work and combine different information expressed using different data types. Consequently, methods that are able to deal with data expressed using different representation models are needed. Thanks to them, data can be expressed in a way that it can be compared and managed together, without having to take into account the origin of the information.

In this kind of scenarios where data is heterogeneous and it is represented using fuzzy sets theory and linguistic modeling, multi-granular fuzzy linguistic methods (multi-granular FLM) [18–21] become essential. Thanks to them, it is possible to carry out conversion operations in order to homogenize the information. In such a way, the system can easily work with all the information. Multi-granular FLM can also allow users to select the LTSs that better fits them. Therefore, user-system communication is improved. In this paper, three new different ways of how multi-granular FLM processes can be applied when fuzzy ontologies are built and managed are proposed and analyzed. To do so, advantages, drawbacks and viability of the different processes depending on the type of information we are dealing with, are presented.

In Section 2, basis needed to understand the proposed methods are introduced. In Section 3, some new methods to solve the multi-granularity treatment problem that is present in FOs are proposed. In Section 4, examples of the exposed approaches described in Section 3 are showed. In Section 5, advantages and drawbacks of the proposed methods are analyzed. Finally, some conclusions are pointed out.

2. Preliminaries

To make this paper as self-contained as possible, this section introduces some concepts and methods to be referred to thorough this paper. In subSection 2.1, multi-granular FLMs are introduced. In subSection 2.2, Fuzzy Ontologies basis are exposed. In subSection 2.3, we describe some features of the Fuzzy Wine Ontology that we use for computing the example results.

2.1. Basis of multi-granular FLM

Linguistic Modeling [17] and the way that it allows people to communicate with computers using words has become an important improvement in human–computer communication. Thanks to it, humans can express themselves using imprecise information as it is the way that they are more used to provide it [22].

Traditional linguistic modeling usually force all the involved users to express themselves using the same LTS. This restriction can become a disadvantage since the selected LTS might not be the best choice for all of them. That is, there can be users that do not feel comfortable with certain LTSs. This situation usually appears when the LTS granularity does not fit the knowledge of the problem that the user has. Therefore, if the user has a wide knowledge of the dealt issue, he/she would prefer to use an LTS that have a high granularity. Then, user can provide more precise information to the system. On the other hand, if a user that does not have too much knowledge about the problem is given a set of words too big for him/her, then the user would get lost among all the possibilities that he/she is given. Consequently, he/she would have problems to provide the required information. In order to solve this kind of situations, it is mandatory that users are allowed to work with LTSs that are specifically designed for them.

In order to solve this problem, multi-granular fuzzy linguistic modeling [23,24] can be used. Thanks to multi-granular FLM, users that utilize the same computer system can provide information with the LTS that better fits them. Thus, user confidence and expressibility are increased and the provided information becomes more accurate and reliable. The usual process followed by multi-granular FLM methods in order to deal with different LTSs is showed below:

1. **Providing preferences**: Users provide the required information using the LTS that they prefer.
2. **Information uniformization**: All the information expressed using different sources is transformed into words expressed by the same LTS. This LTS is usually called the basic LTS (BLTS).
3. **Carrying out computations**: Once that all the information has been uniformed, it is possible for computers to carry out the required computations.

In Fig. 1, three LTSs are defined over the same space range. Vertical lines stablish correlations among them and can be used to define multi-granular transformation functions.

In the recent literature, there are several multi-granular FLMs methods. For instance, in [26], discrete fuzzy numbers are used in order to design a multi-granular FLM method. No membership functions are necessary in order to carry out the required operations. That is, all the computations are made using discrete fuzzy numbers environment. In [27], qualitative description spaces are used in order to carry out the required linguistic labels transformations. Distances in the space of qualitative assessments are used to carry out the required transformations. In [28], a multi-granular FLM method for dealing with multi-granularity uncertain linguistic group decision making problems with incomplete weight information is presented. In order to carry out computations, triangular fuzzy numbers are used. Operations are carried out using the membership function of the fuzzy numbers. In [29], a normalized numerical scaling method that is able to determine semantics of linguistic labels that belong to different LTSs is presented. This method works with either balanced or unbalanced LTSs. In [30], a multi-granular FLM for unbalanced LTSs is defined. For carrying out computations, linguistic distribution assessments with exact symbolic proportions are used. Aggregation and transforming operators are defined over this environment.

There are also several papers in the recent literature that applies multi-granular FLM methods to solve problems. For instance, in [31], multi-granular FLM methods are used to create a Project evaluation method. Non-formatted text information is used in the process. In [32,33], multi-granular FLM methods are applied to create a consensus based group decision making method.

In this paper, when transformations among labels from different LTSs want to be carried out, the multi-granular FLM exposed in [25] is used. This method is based on the concept of linguistic hierarchies (LHs) and the 2-tuple ordinal fuzzy linguistic modeling [34]. A linguistic 2-tuple is defined as a tuple (s, α) where s is an ordinal linguistic label and $\alpha \in [-0.5, 0.5)$ is called the symbolic