Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems 285 (2016) 1–5

FUZZY sets and systems

www.elsevier.com/locate/fss

Editorial

Fuzzy sets and systems + natural language generation: A step forward in the linguistic description of time series

The automatic generation of linguistic descriptions of data is a relatively recent research area originated in the Natural Language Generation field, which is the area of Computational Linguistics that is concerned with the mapping from non-linguistic to linguistic expressions [1]. The objective is to develop computational *data-to-text systems*, able to generate texts from non-linguistic (usually numerical) input data. Such texts, that we call *linguistic descriptions of data*, express knowledge extracted from the data using natural language.

Despite being a relatively young research area, data-to-text systems have reached the maturity not only in terms of technical achievements, but of commercial success, with several existing companies and technology transfer centers developing systems which are being used both by private and public organizations. Most of these success cases, and a large amount of technical developments, have focused on time series data, i.e., in a broad sense, sequences of data regarding a given phenomenon that are ordered in time.

In parallel to the efforts of researchers in the Natural Language Generation area, but mostly unaware of the relation to data-to-text systems, the generation of linguistic descriptions of data has been also dealt with by researchers in the Fuzzy Sets and Systems community, mostly under the name of *linguistic summarization*. Beginning with the work by R.R. Yager [7], fuzzy quantified statements and other closely related tools were employed, mostly for the purpose of flexible querying in databases. In recent years, the research in this area is more and more in confluence with the view and objectives of the Natural Language Generation community.

Our initiative is motivated by three main goals:

- To increase awareness of the area of generating linguistic descriptions of data, their objectives, techniques and open problems, as well as its relevance to both the Natural Language Generation and Fuzzy Sets and Systems research communities.
- To show some of the most recent research and development advances in generating linguistic descriptions of data. We have restricted our interest to time series data as the most usual kind of data in the field, with particular issues, having at the same time a more focused topic.
- To stress that the mentioned research areas have complementary capabilities for, and to promote joint collaboration as a way to make a significant step forward in, the development of data-to-text systems.

The contributions include both methodological proposals and applications, as well as two works aimed at analyzing the problem from a general point of view, considering both Fuzzy Set and Systems and Natural Language Generation. Let us first present these two research works:

- *On generating linguistic descriptions of time series*, by N. Marín and D. Sánchez, provides a general approach to the process of generating linguistic descriptions of time series data, and a global view of the concepts, elements and processes involved, as well as open problems. This general approach unifies the current state of the art both from the Natural Language Generation and Fuzzy Sets and Systems communities, as it is shown in the paper by

http://dx.doi.org/10.1016/j.fss.2015.12.003

studying the correspondence between the components of the most significant existing proposals, and the elements of the proposed global view.
In this general approach, the generation process consists of two main tasks: a *knowledge extraction* process which in a broad sense can be considered as KDD (Knowledge Discovery in Databases), and a *linguistic expression process* which enhances the understandability and usefulness of the obtained knowledge by appropriately expressing it using natural language. These tasks have at their basis three pillars of the generation problem: a *knowledge representation formalism*, an *expression language*, and a *quality framework*. On the basis of these three aspects, the main tools and techniques that can be used in the generation of linguistic descriptions of time series are analyzed, deeply revising the main contributions existing in the literature, from the pioneering work of Karen Kukich [8] to the more recent proposals.
As another key contribution, the paper shows how in both the areas of KDD and Natural Language Generation it is widely recognized the potential contributions of uncertainty representation techniques, particularly those related to the Fuzzy Set Theory and extensions. In KDD and Machine Learning, Fuzzy Set Theory has the potential to produce models that are more comprehensible, less complex, and more robust, being especially useful for representing "vague" patterns and modeling and processing various forms of uncertain and incomplete information [2]. In Natural Language Generation, Fuzzy Set Theory plays a key role in filling the semantic gap between data and linguistic terms and expressions, and in dealing with the different kinds of uncertainty inherent to the linguistic expression of knowledge about real world data [3–5], in line with the well-known suitability and high potential of fuzzy sets for representing the semantics of natural language expressions [6].

- In the work entitled *On the role of linguistic descriptions of data in the building of natural language generation systems*, A. Ramos-Soto, A. Bugarín and S. Barro present a review of the state of the art in the fields of Natural Language Generation and Fuzzy Sets and Systems (linguistic description of data) as two complementary areas of research. The paper describes the definitions and basic concepts of the two areas and reviews a variety of representative systems, use cases and real applications that can be found in the literature.
  The authors state in their work that the use of linguistic description of data together with natural language generation systems is a viable alternative to solve real problems of textual information generation: while Natural Language Generation can bring its proven ability to produce more sophisticated descriptive texts, the area of Linguistic Description of Data can provide tools for extracting and representing knowledge contained in the data and manage the inherent imprecision of natural language through results of the Fuzzy Set Theory.

In addition, this issue contains five remarkable works with methodological contributions:

- V. Novák brings a work entitled *Linguistic characterization of time series* which describes how natural language sentences with information of time series data can be generated using two methods of Soft Computing such as F-transform and Natural Fuzzy Logic. The work can be split in two parts. In the first, the author provides an overview of these two methods and how they have been applied in his previous work to describe and predict the trend of a data series. Then, a second part of the paper is aimed to outline possible future research regarding the mining of linguistic information from time series. In this part, the author suggests an algorithm for finding intervals with monotonous behavior in the series and provides some interesting ideas about the extraction of linguistic summaries of the characteristics of the series through the use of the Theory of Intermediate Quantifiers and generalized Aristotle's syllogisms.
- This special issue also counts on a work where tools of Fuzzy Sets Theory are applied to manage uncertainty in temporal expressions. In Natural Language Generation systems, temporal uncertainty in raw data can complicate the inference of temporal and causal relationships between events and influence in the quality of the generated texts. In the paper by A. Gatt and F. Portet (*Multilingual generation of uncertain temporal expressions from data: a study of a possibilistic formalism and its consistency with human subjective evaluations*), the authors introduce a framework to represent and reason with temporal uncertainty based on Possibility Theory and propose a model that uses the outcomes of such temporal reasoning to select linguistic expressions to convey uncertainty to the reader. Authors remark in the paper that uncertainty should be communicated to the end user.
  The model used by the authors is based on Fuzzy Temporal Constraint Networks (FTCN) [9]. They work on experimental data from three languages and test the correlation between their predictions and human subjective uncertainty in different scenarios. The research reported is an interesting novel point of view that understands