Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems 285 (2016) 118–139

FUZZY sets and systems

www.elsevier.com/locate/fss

Description of multivariate time series by means of trends characterization in the fuzzy domain

Juan Moreno-Garcia [a,*], Javier Abián-Vicén [b], Luis Jimenez-Linares [c], Luis Rodriguez-Benitez [c]

[a] *Escuela de Ingeniería Industrial, Universidad de Castilla La Mancha, Avd. de Carlos III, 45071, Toledo, Spain*
[b] *Facultad de Ciencias del Deporte, Universidad de Castilla La Mancha, Avd. de Carlos III, 45071, Toledo, Spain*
[c] *Escuela Superior de Informática, Universidad de Castilla La Mancha, Paseo de la Universidad s/n, 13071, Ciudad Real, Spain*

Received 28 August 2014; received in revised form 19 March 2015; accepted 25 May 2015

Available online 9 June 2015

Abstract

This paper presents a new method to automatically obtain linguistic descriptions from a multivariate time series. The first step consists of the induction from the multivariate time series of a fuzzy model called the temporal fuzzy model. Later this fuzzy model is analyzed in order to obtain trends based on the output variable. With this information, the trends for the input variables are also obtained and together with the output variable trends are represented in such a way that the temporal evolution of the trend is stored within the own-trends. In a trend there are always some elements or points of interest that are more relevant than others for generating a relevant final description. These elements are also extracted and stored, making the final description generation process more efficient. An event search process generates the final linguistic description. The events are identified in the trends or in the structure containing the points of interest previously selected. Once an event is identified, new text is added at the end of the linguistic description. The design of the events and the text related to each event is made with the cooperation of experts in the field of application. The presented approach is checked in sports, more concretely, in countermovement jumping.

Keywords: Multivariate time series; Linguistic description; Temporal fuzzy models; Countermovement jump

1. Introduction

Time series can be found in a great number of different systems and they are classified either as univariate (TS) or multivariate time series (MTS). The latter type of series contains more than one variable and in such a case the relations over time between the variables can be studied [1]. In the research in this field, great efforts have been made to directly extract text from the time series. Usually, the relevant information is obtained using a knowledge acquisition

* Corresponding author.
E-mail address: juan.moreno@uclm.es (J. Moreno-Garcia).
URL: http://www.uclm.es (J. Moreno-Garcia).

http://dx.doi.org/10.1016/j.fss.2015.05.011
0165-0114/

process. The knowledge extracted from the data is then represented as a model, making the real system represented by the time series more comprehensible.

In this paper, the "linguistic description process" is understood as a method to automatically synthesize a text. This text must describe the most interesting knowledge obtained from the numerical data. This process can be carried out by experts, but the great amount of data renders their work difficult, and then the automation of this process is interesting for efficiently generating such descriptions. The quality of the descriptions obtained should be similar to that of the descriptions generated by experts, and for this reason an expert would be needed in some situations to supervise the process. For instance, classical areas of application of these methods include clinic monitoring [2], sports [3], human gait [4], weather and contamination [5,6], traffic [7], etc.

Moreover, the use of linguistic variables as proposed by Zadeh [8–10] makes it possible to achieve representations that can be considered closer to human reasoning than other ones. Also, fuzzy logic [11] helps treat the uncertainty and imprecision that can be present in a TS or an MTS. Usually, fuzzy logic makes use of fuzzy models to represent a system, with these fuzzy models usually being a simple collection of fuzzy rules [12,13]. Often these models are automatically induced [14]. Fuzzy logic has been applied to obtain linguistic descriptions with reasonable success. For example, there are some interesting publications that try to generate linguistic descriptions from the human gait [15]. Alvarez-Alvarez and Triviño [16,17] use Granular Linguistic Models of Phenomena (GLMP) and Fuzzy Finite State Machines to create a linguistic model of the human gait. With the obtained model, they construct a human linguistic description of this phenomenon, focused on the assessment of the gait quality. Also, a practical application is implemented to analyze the gait quality of healthy individuals and people with lesions in their limbs. The research group headed by Triviño has linguistic description as one of its prominent research lines. They describe different types of systems like human gait [17], traffic [18], circular structures of the surface of Mars [19], or Astronomy [20]. All these papers have used GLMP [17,21] as a tool to achieve a linguistic description of the studied system. Moreno-Garcia et al. [4] also used fuzzy logic to model human gait [4]. They presented a method to automatically generate a linguistic model using an MTS as input. The method is based on CART [22] and ID3 [23]. That approach has been applied to obtain a model of the human gait with the aim of describing every phase of the human gait by means of the rules of the obtained model.

On the other hand, the aim of this paper is to create a linguistic description of an MTS using a representation based on fuzzy logic that analyzes the trends on the input and the output variable. The use of trends shows the MTS variable evolution and the relation among them. It can be considered as a way to provide a coarser informed analysis of the MTS. For this reason, we consider that trends are a good approach to generate linguistic descriptions. The presented method needs the support of an expert when the structure of the linguistic descriptions is defined.

Finally, the experimentation is focused on sports: more concretely, on the description of a countermovement jump (CMJ). It must be said that jumping is one of the most common activities in physical education and in many sports, and it is frequently used as a physical aptitude test and as a method of evaluation of the explosive force in the lower extremities [24,25]. In recent years, the countermovement jump test has gained a lot of attention because of the possibility of determining the leg contribution and contractile characteristics of individuals [26] with a jump performed on a force platform: something which is simple to do and easy to analyze. Countermovement jumps have shown very high reliability values, from 4.3% to 6.3% [27,28] and have been found to be the most reliable and valid field tests for the estimation of the explosive power of the lower limbs in physically active men [26]. We consider it an interesting field in which to test our approach.

This paper is organized as follows. Section 2 presents a brief overview of the method used to induce a temporal fuzzy model (TFM) [29] to linguistically represent an MTS. Section 3 describes the method to generate the linguistic descriptions of the input MTS. Section 4 tests the proposed method, presents a discussion of the results achieved, and finally Section 5 details the main contributions of the proposal as the conclusions of this paper.

2. Induction of a temporal fuzzy model from a multivariate time series

The concept of a temporal fuzzy model was first presented in [29] and a brief description of it will now be given in detail. First of all, the concepts of an MTS and of an ordered set of labels will be defined.

Definition 1. A multivariate time series (MTS) is an ordered sequence of examples $E = \{e_1, e_2, \ldots, e_n\}$, where i represents the instant when the example e_i is captured and n indicates the number of observations [1,29]. Every