Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems 285 (2016) 94–117

FUZZY sets and systems

www.elsevier.com/locate/fss

Linguistic summaries of locally periodic time series

Gilles Moyse [a,b,*], Marie-Jeanne Lesot [a,b]

[a] *Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6, F-75005, Paris, France*
[b] *CNRS, UMR 7606, LIP6, F-75005, Paris, France*

Received 23 September 2014; received in revised form 5 June 2015; accepted 12 June 2015

Available online 26 June 2015

Abstract

This paper proposes a method to linguistically summarise the local periodic components of a time series: it identifies subparts of the data which are periodic, together with their periodicity degree and period, and provides a linguistic description thereof. The generated sentences can be illustrated by the example "Approximately from March to June, the series is highly periodic with a period of exactly 2 weeks". The method proposed to identify local periodic zones relies on the determination of relevant auto-adaptive windows, based on an analytical expression of the probability distribution of the considered periodicity criterion. The linguistic description generation, in the protoform approach framework, expresses three core aspects of the identified periodic intervals, namely their time context or localisation in time, their periodicity and their period. Intensive experiments performed on both artificial and real data validate the proposed method.

Keywords: Local periodicity; Contextualisation; Linguistic summary; Time series

1. Introduction

Linguistic summaries offer a compact and user-friendly way of representing large amounts of data [1–3]. In the case where the considered data are time series, the linguistic summarisation task brings specific questions due to their temporal nature. Furthermore, the properties of such series are often changing over time, which implies that the knowledge extracted from a series at a certain time may be wrong at the next one, and hence needs to be contextualised.

Among the various information conveyed in a time series, periodicity is an important one, frequently used in fields as diverse as astronomy [4], physics [5], energy production [6], speech analysis [7], zoology [8] or biology [9]. Moreover, the question of local periodicity occurs in many applicative contexts [10–13].

This paper proposes a method allowing to simultaneously address the two issues of contextualisation and periodicity: it aims at identifying local periodicity, defined as the occurrence in the time series of periodic patterns located in a specific temporal zone.

* Corresponding author at: Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6, F-75005, Paris, France
E-mail address: Gilles.Moyse@lip6.fr (G. Moyse).

http://dx.doi.org/10.1016/j.fss.2015.06.016
0165-0114/

The method proposed in this paper is called Local Detection of Periodic Events (LDPE). It detects and characterises the periodic zones in terms of periodicity and period, locates them in time and renders this local periodic information into human readable sentences. It is based on the Detection of Periodic Events (DPE) method [14], and brings two important improvements: an auto-adaptive window, based on a probability distribution allowing to automatically adapt a periodicity computation window to the data, and a set of new linguistic rendering features to generate sentences such as "Approximately from March to June, the series is highly periodic with a period of exactly 2 weeks".

In Section 2, existing works related to both the issues of linguistic summarisation and periodicity detection are presented. The DPE method is detailed in Section 3. The different steps of periodicity contextualisation of LDPE, defined as a generalisation of DPE, are presented in Section 4. Section 5 is dedicated to the linguistic rendering issue raised by such local periodic events and presents formulations to enrich existing models and, in particular, to add new time localisation features. The experiments both on artificial and real data and the obtained results that validate our approach are presented in Section 6. Finally, a conclusion to this work and future works are given in Section 7.

2. Related works

The question of detecting periodic parts of a times series lies at the crossroad of several domains, namely linguistic data summarisation and periodicity detection. Some of the existing works in each of these two domains are successively presented in the next subsections.

2.1. Linguistic summarisation

Two sides of linguistic summarisation are presented in this subsection: first, the general summaries and second, the temporal ones.

General summaries Linguistic summaries aim at building compact representations of datasets, in the form of natural language sentences describing their main characteristics. Two principal approaches exist to this aim [15,16], one using fuzzy protoforms and one based on natural language generation (NLG).

Fuzzy Linguistic Summaries (FLS), introduced in the seminal papers [1,2,17], are built on "protoforms" whose basic form is defined as "QX are A" where Q is a quantifier (e.g. "most" or "around 10"), A a linguistic modality associated with one of the attributes (e.g. "young" for the attribute "age") and X the data to summarise. Numerous criteria to evaluate the relevance of a candidate protoform instantiation have been proposed. The first and maybe most important one is the truth degree, proposed in the seminal paper [1]: it measures the extent to which the data coincides with the considered summary, based on the Σ-count of the dataset according to the chosen fuzzy modality.

Other criteria include the degree of focus [18], which measures the support of a given attribute in the database, and the degree of appropriateness [19] which quantifies the extent to which a summary is "surprising". Compound measures combine several criteria such as fulfilment, relevance, length, coverage, specificity, compatibility and non-ambiguity of a summary [20] or its coverage, brevity, specificity and accuracy [21].

In the NLG framework, several approaches have been developed as well, more focused on sentences generation than on data extraction. Indeed, such systems, as EasyText performing polls analysis [22], are based on a user-defined set of rules used against a database to generate the result sentences. Since the method proposed in this paper is more oriented toward data analysis, NLG techniques are not further investigated.

Temporal summaries FLS coping with the temporal dimension of time series have also been considered, as for instance specific protoforms expressing duration or time localisation of specific events: they can e.g. be based on fuzzy temporal protoform grammar [23] or on a specific hierarchical time scale [24].

Additionally, the extraction of temporal features from time series, such as trends, in order to summarise them later using standard FLS, has also been proposed [25]. Another kind of extracted temporal feature such as the exceptional character of an event in time compared to a reference value is detected and included in a FLS in [26].

The DPE method [14] and its variants [27] focus on the issue of periodicity, not taken into account in the previous approaches. They propose linguistic summaries expressing this periodicity as well as the period of a time series, in a human friendly way. The DPE method, on which the approach proposed in this paper relies, is presented in further details in Section 3.