Applied Soft Computing 40 (2016) 441–454

Contents lists available at ScienceDirect

Applied Soft Computing

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

Change points detection in crime-related time series: An on-line fuzzy approach based on a shape space representation

Fabrizio Albertetti [a,*], Lionel Grossrieder [b], Olivier Ribaux [b], Kilian Stoffel [a]

[a] Information Management Institute, University of Neuchatel, CH-2000 Neuchatel, Switzerland
[b] School of Criminal Sciences, University of Lausanne, CH-1015 Lausanne-Dorigny, Switzerland

ARTICLE INFO

Article history:
Received 20 April 2015
Received in revised form 8 October 2015
Accepted 2 December 2015
Available online 17 December 2015

Keywords:
Change points detection
Qualitative description of data
Time series analysis
Fuzzy logic
Crime analysis

ABSTRACT

The extension of traditional data mining methods to time series has been effectively applied to a wide range of domains such as finance, econometrics, biology, security, and medicine. Many existing mining methods deal with the task of change points detection, but very few provide a flexible approach. Querying specific change points with linguistic variables is particularly useful in crime analysis, where intuitive, understandable, and appropriate detection of changes can significantly improve the allocation of resources for timely and concise operations. In this paper, we propose an on-line method for detecting and querying change points in crime-related time series with the use of a meaningful representation and a fuzzy inference system. Change points detection is based on a shape space representation, and linguistic terms describing geometric properties of the change points are used to express queries, offering the advantage of intuitiveness and flexibility. An empirical evaluation is first conducted on a crime data set to confirm the validity of the proposed method and then on a financial data set to test its general applicability. A comparison to a similar change-point detection algorithm and a sensitivity analysis are also conducted. Results show that the method is able to accurately detect change points at very low computational costs. More broadly, the detection of specific change points within time series of virtually any domain is made more intuitive and more understandable, even for experts not related to data mining.

1. Introduction

The analysis of time series naturally arises in crime analysis as well as in any data-driven domain. Finding sudden changes in criminal activities is a particular task known as change points detection. In this paper, a flexible on-line change points detection method for helping crime analysts to easily and understandably monitor changes is proposed. Change points are detected in two steps: the segmentation of the time series and the querying of points with a fuzzy inference system.

1.1. Motivation

Knowledge extraction of time series can be viewed as an extension of traditional mining methods with an emphasis on the temporal aspect. Among these, *Change points detection* methods focus on finding time points at which data *suddenly* change (in contrast to *slow* changes). Many studies have shown interesting applications of change points detection in various domains. These methods are based on neural networks, regressions, or other statistical models, with an emphasis on the efficiency of these methods. However, only a few consider approaches with these two properties: a meaningful and expressive subspace representation of the time series, and a dynamic segmentation process without fixed-sized windows, linked together in a flexible way.

In the domain of crime analysis, such flexible and intuitive approaches for change points detection are particularly sought, especially for crime trends monitoring. Previous studies from the authors [1–4] emphasize on the usefulness of crime trends monitoring activities and advocate the use of appropriate methods for considering the specificities and the constraints of the crime analysis domain, that is basically dealing with uncertainties. The automated process of change points detection is considered as a major step in the production of intelligence, supporting the activity of crime analysis (also sometimes referred to crime intelligence).

Flexible change points detection methods are critical for supporting analysts in their daily tasks, especially for the monitoring of serial and high-volume crimes (e.g., burglaries). Most of the time, crime analysts have no particular background in time series analysis, but still need to analyze and monitor crime trends. These trends are drawn into the whole activities and are not always perceived by police forces. As for example, querying criminal activities about a particular increase in crime trends for targeted police interventions, as well as querying patterns of changes for the general understanding of crime phenomena are common tasks.

Finding changes in crime trends assumes two conditions: (a) the actual existence of a trend, and (b) its detection within the data. The first condition is far from obvious, but as crime analysis is founded on environmental criminology theories, a justification for the existence of crime trends appears [4–6]. The second is generally simply assumed, but difficult to detect in massive data sets and needs intuitive and understandable analytical methods.

Although the proposed method is a specific answer for the domain of crime analysis, we believe that it has great potential applications in several domains. As an example, in the financial domain, it proves very useful to find and query change

* Corresponding author. Tel.: +41 32 718 14 49.
E-mail addresses: fabrizio.albertetti@unine.ch (F. Albertetti), lionel.grossrieder@unil.ch (L. Grossrieder), olivier.ribaux@unil.ch (O. Ribaux), kilian.stoffel@unine.ch (K. Stoffel).

http://dx.doi.org/10.1016/j.asoc.2015.12.004
1568-4946/

points in real-time, giving the investors a flexible means to detect trends indicating the right moment for selling or buying stocks.

1.2. Contribution of this paper

The method proposed in this paper, for the Fuzzy Change Points Detection in crime-related time series (FCPD), aims to focus on flexibility and intuitiveness. To achieve this purpose, the method combines a segmentation step and a querying step. Moreover, the following characteristics make FCPD unique:

- a meaningful and expressive representation of the time series is used;
- the segmentation is *dynamic*, that is, segments are set according to the underlying shapes of the time series, without using a fixed-size window parameter;
- changes are queried with linguistic terms, using a fuzzy inference system;
- the method does not rely on training sets;
- the method is on-line and iterative, i.e., change points can be detected with past values only and there is no need to compute the entire model at each new observed value. The computational cost is very low; it is related to the size of the approximating polynomials (instead of the number of the observations).

Indeed, with the use of a meaningful representation and a dynamic segmentation, change points can be more easily described and identified. The segments found in a time series, reflecting change points, are described with meaningful estimators such as the average, the slope, the curvature, etc. Then, with the use of a fuzzy inference system, a query can be specified using linguistic terms describing the geometric estimators. It becomes then easier, for instance, to query a time series about the most abrupt changes in terms of slope. In the example "IF *average* is *low* AND *slope* is *very_high*, THEN *pertinence* is HIGH", the inference system would return a high score on segments representing shapes of the given description. This approach makes the querying of change points particularly intuitive and flexible, especially for domain experts.

1.3. Structure of this paper

The remainder of this paper is structured as follows: in Section 2, a literature review in the mining of time series is provided; Section 3 introduces some concepts in the preparation, representation, and analysis of time series; Section 4 details FCPD, a step-by-step method for the fuzzy querying and detection of change points in crime-related time series; an empirical validation on synthetic and real-world data is conducted in Section 5; results are discussed in Section 6; and finally in Section 7 a conclusion is drawn from the experiments and some tracks for future work are suggested.

2. Literature review

Change points detection has numerous application domains, as for example finance, biology, ocean engineering, medicine, and crime analysis. It is considered as a final objective in the whole process of time series analysis amongst classification, rules discovery, prediction, and summarization. Almost all of these mining tasks require data preparation, namely the representation of the time series, its indexing, its segmentation, and/or its visualization. In this section, we propose a review of these steps, before comparing existing methods for change points detection. An extensive review of the analysis of time series can be found in [7], as well as a general methodology in [8].

2.1. Representation of time series

Many representation models of time series have been dealt with in the literature, each claimed with relative advantages and drawbacks. Two main categories are symbolic representations and numeric representations. Symbolic representations are less sensitive to noise and are usually computationally faster. For the last decade, the community has been paying particular attention to the Symbolic Aggregate Approximation (SAX) representation [9,10], with the main advantages to reduce the original dimensionality of the data, being on-line, and having a robust distance measure. However, it does not cover all needs. In [11], a numeric representation – which differs from SAX and many others by giving a meaning to the representation – is used to perform several mining tasks. This *shape space* representation uses coefficients as shape estimators of the time series it represents, leading to an intuitive description.

2.2. Segmentation of time series

Most mining methods use subsequences (or segments) of time series as input to the analysis. Segmentation algorithms with the approach of a sliding window are simple to use but present the main drawback of being static, i.e., segmenting the time series according to a fixed and exogenous parameter (e.g., the length of the window) without considering the observed values. Other algorithms, based on a bottom-up or a top-down approach are considered as dynamic (e.g., by using some error criteria as segmentation thresholds) but need the whole data set to operate. These off-line algorithms usually perform better in terms of accuracy but have higher computational costs and are not suitable for real-time applications. A combination of the aforementioned algorithms, namely the SWAB segmentation algorithm, is presented in [12]. A study [13] provides benchmarks on these claims and as a result suggests that SWAB is empirically superior to all other algorithms discussed in the literature. As we believe there is no silver bullet, each application has its own requirements. A more flexible approach is the SwiftSeg algorithm [14], providing a dynamic and on-line approach to segmentation, with the possibility of a mix between growing and sliding window. Another interesting segmentation approach [15], specific to stock mining and described as dynamic, is based on the identification of perceptually important points (PIP).

2.3. Fuzzy analysis of time series

A small subset of temporal mining methods takes advantage of the characteristics offered by fuzzy logic and fuzzy sets. The concept of fuzzy time series has first been defined by Song and Chissom in [16,17], with an application in class enrollment forecasting. Soon followed multiple variations and improvements of the basic method (e.g., [18–20], or [21]), with their own types of fuzzy inference systems (FIS). Two common FIS, namely the Mamdani inference system [22] and the Takagi-Sugeno inference system [23], can be intuitively used to deal with uncertain and flexible data. In [24], an application in finance uses an FIS for pattern discovery. In [25], prediction of long shore sediments is also dealt with the use of an FIS. In parallel, a combination of FISs and neural networks have found an origin in [26]. As for examples, the prediction of time series is performed with dynamic evolving neuro-fuzzy inference systems [27], the classification of electroencephalograms [28], as well as the prediction of hydrological time series [29].

2.4. Change points detection

Change points detection in time series analysis has been thoroughly investigated, mainly using statistical models (see [30] for a general introduction). Reeves et al. [31] attempt to review and compare the major change points detection methods for climate data series.

More specifically, related approaches for change points detection have been investigated in a relatively limited set of studies. For example, a statistical based approach using fuzzy clustering is described in [32,33]. Verbesselt et al., in [34,35], detect breaks for additive seasonal and trends (BFAST), with a principal application is phenology. To deal with imprecise observation in time series, changes are analyzed with fuzzy variables in [36]. In [37], a contextual change detection algorithm addresses relative changes with respect to a group of time series. In [38] and [39], the utility of a framework for outliers detection of time series prediction is highlighted. In [21], the need to use linguistic values for comprehensible results is advocated, where fuzzy time series mining is used for association rules between data points (but not between segments) with fixed-size window. A qualitative description of multivariate time series with the use of fuzzy logic is presented in [40]. Yu et al.