Expert Systems with Applications 42 (2015) 7664–7675

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Mixed dissimilarity measure for piecewise linear approximation based time series applications

Zoltán Bankó *, János Abonyi [1]

University of Pannonia, Department of Process Engineering, P.O. Box 158, H-8200 Veszprém, Hungary

ARTICLE INFO

Article history:
Available online 10 June 2015

Keywords:
Dynamic time warping
Local distance
Piecewise linear approximation
Similarity
Constraining

ABSTRACT

In recent years, expert systems built around time series-based methods have been enthusiastically adopted in engineering applications, thanks to their ease of use and effectiveness. This effectiveness depends on how precisely the raw data can be approximated and how precisely these approximations can be compared. When performance of a time series-based system needs to be improved, it is desirable to consider other time series representations and comparison methods. The approximation, however, is often generated by a non-replaceable element and eventually the only way to find a more advanced comparison method is either by creating a new dissimilarity measure or by improving the existing one further.

In this paper, it is shown how a mixture of different comparison methods can be utilized to improve the effectiveness of a system without modifying the time series representation itself. For this purpose, a novel, mixed comparison method is presented for the widely used piecewise linear approximation (PLA), called mixed dissimilarity measure for PLA (MDPLA). It combines one of the most popular dissimilarity measure that utilizes the means of PLA segments and the authors' previously presented approach that replaces the mean of a segment with its slope.

On the basis of empirical studies three advantages of such combined dissimilarity measures are presented. First, it is shown that the mixture ensures that MDPLA outperforms the most popular dissimilarity measures created for PLA segments. Moreover, in many cases, MDPLA provides results that makes the application of dynamic time warping (DTW) unnecessary, yielding improvement not only in accuracy but also in speed. Finally, it is demonstrated that a mixed measure, such as MDPLA, shortens the warping path of DTW and thus helps to avoid pathological warpings, i.e. the unwanted alignments of DTW. This way, DTW can be applied without penalizing or constraining the warping path itself while the chance of the unwanted alignments are significantly lowered.

1. Introduction

In the last decades, one of the most important trend of expert systems has been the emerging interest in data-driven approaches. The databases are not seen as the simple input of a hard-coded decision system anymore, instead, they are also actively used to improve the system itself by utilizing data mining.

The availability of easily accessible computational resources and storage capacities has made it possible to apply classical data mining techniques on time series data. Such methods have rapidly gained attention in engineering applications, including process engineering (Singhal & Seborg, 2005), medicine (Tormene, Giorgino, Quaglini, & Stefanelli, 2009), bioinformatics (Aach & Church, 2001), chemistry (Abonyi, Feil, Németh, & Árva, 2005), finance (Rada, 2008), biometrics (Gavrila & Davis, 1995; Kholmatov & Yanikoglu, 2004; Vajna, 2000) and even tornado prediction (McGovern, Rosendahl, Brown, & Droegemeier, 2011) and sewer flow monitoring (Dürrenmatt, Giudice, & Rieckermann, 2013), because of the effective and efficient methods available for solving time series-related problems. But these methods are not limited exclusively to temporal data. Every task that can be transformed to a sequence-related problem is also a candidate, such as motion detection (e.g. gun draw detection (Ratanamahatana & Keogh, 2004)), handwritten word recognition (Rath & Manmatha, 2003), off-line signature verification (Piyush Shanker & Rajagopalan, 2007), detection of similar shapes in a large image dataset (Xi, Keogh, Wei, & Mafra-Neto, 2007), etc.

* Corresponding author. Tel.: +36 88 624209.
E-mail addresses: bankoz@tvn.hu (Z. Bankó), abonyij@fmt.uni-pannon.hu (J. Abonyi).
[1] Tel.: +36 88 624209.

http://dx.doi.org/10.1016/j.eswa.2015.06.003
0957-4174/

The increasing use of time series data has initiated a great deal of research and development attempts in the filed of data mining (Fu, 2011) and thus a continuously growing knowledge base has been built up to utilize time series data even better. Two things are, however, common in every time series-based data mining task: a representation of the temporal data has to be selected and a dissimilarity measure has to be defined between the represented values. The representation thus not only determine the tightness of the approximation but also the dissimilarity measure to use and, in the end, the result itself. Motivated by this fact, large number of time series representations — alongside the matching dissimilarity measures — have been introduced by the time series data mining community: episode segmentation (Cheung & Stephanopoulos, 1990), discrete Fourier transform and discrete wavelet transform-based approaches (Agrawal, Faloutsos, & Swami, 1993; Chan & Fu, 1999), Chebyshev polynomials (Cai & Ng, 2004), symbolic aggregate approximation (Lin, Keogh, Lonardi, & Chiu, 2003), perceptually important points (Fu, Chung, Luk, & Ng, 2008), derivative time series segment approximation (Gullo, Ponti, Tagarelli, & Greco, 2009), implicit polynomial curve-based approximation (Wu & Yang, 2013), etc. All proposed to compete and replace classical time series representations such as piecewise linear approximation (PLA) that has been used — and still preferred — in several engineering applications thanks to its intuitiveness and ease of visualization. Data acquisition systems also echoed this trend and in most systems different approximations of the raw data can be selected allowing to choose the one, which suits most the performed task.

In many cases, however, the representation cannot be changed as it is provided by an unchangeable hardware device or software component. Cars developed according to the AUTOSAR concept (www.autosar.org, 2003) are good examples of that. AUTOSAR is based on the assumption that elements (sensors, processing units, application logic, etc.) of a given automotive functionality made by different suppliers can be easily combined, if the interfaces between the hardware/software elements are strictly defined. It is not an uncommon situation when an analog sensor (the sensing element) and its control unit are provided by two different suppliers. Similarly, the basic software, which operates the control unit, and the application logic can also be created by separate companies. In such a complex and interdependent environment, an analog sensor is often coupled with an application-specific integrated circuit (ASIC) and/or a software library to provide — a usually highly simplified — digital representation of the analog signal. This representation is then forwarded by the control unit through the basic software to the application logic. As an example, Fig. 1 shows the analogue (raw) signal of an automotive ultrasonic sensor and its PLA representation, which can be used by the application logic. Due to the already given representation, the only possibility for the application logic provider to gain more accuracy is not to modify the representation itself but the dissimilarity measure.

Although the creation of mixed dissimilarity measure for PLA (MDPLA), presented in this paper, was motivated by a similar situation —, i.e. PLA segmentation could not be changed —, the development was also driven by the need of minimization of pathological alignments often created by dynamic time warping (DTW[2]).

2. Alignment problem of dynamic time warping

The unwanted alignments are created when the feature considered by the local dissimilarity measure of DTW is similar between a relatively small section of one time series and a much larger section of another time series. This phenomenon can usually be seen when DTW *fail to find obvious, natural alignments in two sequences simply because a feature (e.g., peak, valley, inflection point, plateau etc.) in one sequence is slightly higher or lower than its corresponding feature in the other sequence* (Keogh & Pazzani, 2001).

To avoid such unwanted alignments, global constraints, which limit the warping path how far it can stray from the diagonal, have been introduced by Itakura (1975) and Sakoe and Chiba (1978) (see Fig. 2). According to their experience, *time-axis fluctuation in usual cases never causes too excessive timing difference*, and thus pairing points far away in time would eventually degrade the results.

Limiting the warping path has another advantage, it speeds up the calculation of DTW by a constant factor. As global constraints limit the search space in the warping matrix, local dissimilarity is only necessary to be computed inside the constrained area.

Due to the expected better results and the speed up, most practitioners dealing with DTW utilizing of some form of global constraints independently of whether DTW was used for off-line handwritten signature verification (Güler & Meghdadi, 2008) or for sewer flow monitoring (Dürrenmatt et al., 2013). However, selecting the optimal global constraint — deciding not only which constraint should be used, but whether any constraint should be used at all — is not obvious. When DTW has been used tasks other than speech recognition, *researchers have used a Sakoe-Chiba band with a 10% width for the global constraint* [...] *result of historical inertia inherited from the speech processing community, rather than some remarkable property of this particular constraint* (Ratanamahatana & Keogh, 2005). The width of the Sakoe-Chiba band was optimized later in several applications, e.g., by Long, Fonseca, Foussier, Haakma, and Aarts (2014); however, such an approach can provide a global optimum only and leaves space for local pathological warpings. Thus, to have a better control of the warping path — and to improve DTW results —, different approaches have been utilized.

Ratanamahatana and Keogh (2004) presented the R–K band, which uses a heuristic search algorithm that automatically learns the constraint from the data and locally shrinks the Sakoe-Chiba band, making it narrower. Yu, Yu, Hu, Liu, and Wu (2011) replaced this heuristic search algorithm and utilized large margin criterion to have a better generalization ability on unseen test data. Although the R–K band usually provides better results than the optimized Sakoe-Chiba band, it still suffers from overfitting issues (Niennattrakul & Ratanamahatana, 2009) and in many cases the effectiveness of constrained and unconstrained DTW is the same (Wang et al., 2013). Moreover, Kurbalija, Radovanovic, Geler, and Ivanovic (2014) empirically proved that DTW is very sensitive to the introduction of global constraints and even a small change in an already narrow band changes relation between the time series. Last but not least, application of global constraint can make DTW rigid for real-time applications where there is no chance to do proper preprocessing and/or compensate the initial/ending shifts due to time or hardware limit. On the other hand, it has to be underlined that constraining is a must when fast DTW computation is prioritized, as the most effecting lower bounding functions and indexing methods are based on constraining (Keogh & Ratanamahatana, 2005; Lemire, 2009).

Locally penalized warping can also be used to avoid unwanted warpings (Sun, Lui, & Yau, 2006). In this case penalty is added to every non-diagonal movement of the warping path, i.e. the warping path constrains itself eventually. Instead of using a constant penalty, Clifford et al. (2009) suggested a penalty vector in which each non-diagonal step can have a different cost. A similar approach was presented by Taylor, Zhou, Rouphail, and Porter (2015) who applied the penalty as a scaling factor to the calculated cost. To provide a true data-adaptive constraint, Juhász (2007) used

[2] Review of DTW is omitted for brevity, a comprehensive discussion on it was written by Rabiner & Juang (1993).