Information Sciences 345 (2016) 1–8

Contents lists available at ScienceDirect

Information Sciences

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

Segmenting time series with connected lines under maximum error bound

Huanyu Zhao[a,b,c], Zhaowei Dong[d], Tongliang Li[b,*], Xizhao Wang[f], Chaoyi Pang[b,c,e,**]

[a] SJZ JKSS Technology Co. Ltd, Shijiazhuang, China
[b] Institute of Applied Mathematics, Hebei Academy of Sciences, Shijiazhuang, China
[c] Center for Data Management & Intelligent Computing, Zhejiang University (NIT), China
[d] Radio and TV University, Shijiazhuang, China
[e] RMIT University, Melbourne, Australia
[f] Shenzhen University, Guangzhou, China

ARTICLE INFO

Article history:
Received 12 January 2015
Revised 2 September 2015
Accepted 14 September 2015
Available online 9 October 2015

Keywords:
Piecewise linear
Approximation/representation
Time series data
Haar wavelet
Data compression

ABSTRACT

The error-bounded Piecewise Linear Approximation (PLA) is to approximate the stream data by lines such that the approximation error at each point does not exceed a pre-defined error. In this paper, we focus on the version of PLA problem that generates connected lines in the segmentation for smooth approximation. We provide a new linear-time algorithm for the problem that outperform two of the existing methods with less number of connected segments. Our extensive experiments, on both real and synthetic data sets, indicate that our proposed algorithms are practically efficient.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

A time series is a sequence of data points where each data point is associated with a time stamp. As with most computer science problems, how to efficiently and effectively represent such data is challenging. Essentially, approximate representation is one of the most commonly used methods for data pre-processing and querying. There exist many interesting algorithms or strategies for data approximations, including Fourier Transforms [10], Discrete Wavelet Transform [8], Symbolic Mapping [11], Piecewise Linear Approximation (PLA) [2,5–7] and Piecewise Aggregate Approximation [4].

Recently, the research on maximum-error bound *Piecewise Linear Approximation* (L_∞-bound PLA) has gained some attention. This representation constructs a number of line segments to approximate the stream such that the approximation error on each corresponding point does not exceed a prescribed error bound (L_∞-norm). Xie et al. [12] give an optimal linear-time algorithm[1] that constructs minimum number of line segments in approximation. In their method, the minimum number of line segments is achieved through maximally extending each constructed segment. The general idea of DisConnAlg follows: in order to adjust a line segment to approximate the maximum number of stream points, the algorithm determines the range of all feasible line segments, which is incrementally maintained during the processing of consecutive sequence points. Whenever the current point

[*] Corresponding author.
[**] Corresponding author at: Institute of Applied Mathematics, Hebei Academy of Sciences, Shijiazhuang, China.
E-mail addresses: zhaohuanyu@163.com (H. Zhao), litongliang@tom.com (T. Li), chaoyip@netscape.net (C. Pang).
[1] OptimalPLR algorithm in [12], which is termed as DisConnAlg in this article.

http://dx.doi.org/10.1016/j.ins.2015.09.017
0020-0255/© 2015 Elsevier Inc. All rights reserved.

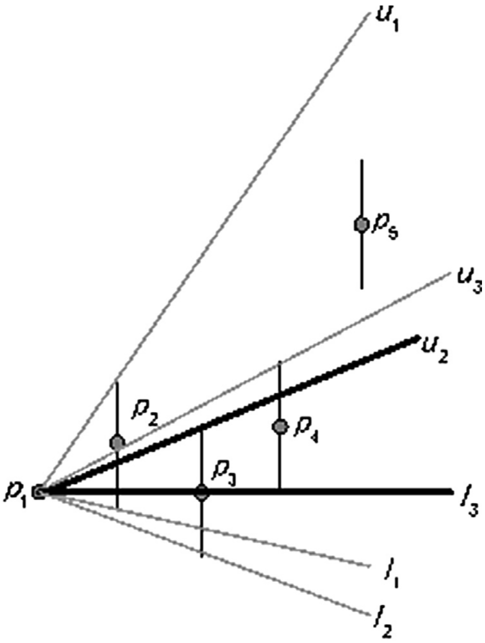

Fig. 1. The process of FSW.

cannot be approximated within the error bound, start a new segment from this point. Furthermore, DisConnAlg can be used to construct connected line segments when used on the restricted feasible space iteratively. That is, constructing the next segment from the feasible space of the last data point of the previous segment.[2] We denote this algorithm that generates connected segments as DConnAlg in this article.

In fact, as an old research problem, there are many algorithms for computing either continuous or discontinuous PLAs under the L_∞ norm, including the original work [14,15] from Bellman and Gluss in the early 1960s. They indicated that this problem can be solved by using dynamic programming method. Other research results on this topic include [17–19]. Paper [17] provided algorithms that only work on special functions of “convex shape”. Paper [18] was about non-connected segmentation. Paper [19] proposed polynomial algorithms. Recently, Liu et al. proposed FSW algorithm [6] that uses the Feasible Space (FS) window method to construct segments from a fixed initial point. Qi et al. [9] extend FS to the polynomial functions in the processing of multidimensional data.

Let $u_i = line(p_1, p_{i+1} + \delta)$ be the line that passes points p_1 and $p_{i+1} + \delta$, and $l_i = line(p_1, p_{i+1} - \delta)$ be the line that passes points p_1 and $p_{i+1} - \delta$. As Fig. 1 shows, Liu’s method first constructs FS to be the area between the lines u_1 and l_1. The feasible space is then incrementally narrowed down to the intersection part of FS and area between of u_i and l_i for the newly arriving points $i + 1$. Continuing the process until the point when FS turns into empty where the next new segment is to be built from this very point iteratively. In the example of Fig. 1, $\{p_1, p_2, p_3, p_4\}$ is approximated by one segment whose FS is the area between u_2 and l_3.

Liu indicated that FSW algorithm outperforms the algorithms of [1,2,13,16] with less number of constructed segments. Liu’s method constructs the FS from the starting point without considering the use of error-tolerant rang $[p_1 - \delta, p_1 + \delta]$ as that of DisConnAlg and DConnAlg. Therefore, the segment constructed by FSW could contain less number of stream points than that of constructed by DisConnAlg or DConnAlg in general. As a result, Liu’s method could output many more segments than that of DisConnAlg or DConnAlg in general.

Our contributions in this article can be summarized as follows:

1. Design and implement ConnSegAlg algorithm. Through incorporating the “Forward-Checking” strategy used in DisConnAlg of [12] and using the “Backward-Checking” strategy, this algorithm has linear time complexity and constructs less number of segments than that of DConnAlg and FSW. Next, we indicate that the number of segments constructed from ConnSegAlg is bounded by $2k - 1$ where k is the optimal number of disconnected segments constructed by DisConnAlg. However, this bound does not hold for DConnAlg and FSW. We indicate that the number of segments constructed by DConnAlg and FSW can be above $2k - 1$ in some situations. Lastly, we show that the $2k - 1$ bound is tight. That is, there exists a stream such that the number of segments constructed from ConnSegAlg equals to $2k - 1$.

[2] Refer to Section 6.3.1 of [12] for details.

Download English Version:

https://daneshyari.com/en/article/392624

Download Persian Version:

https://daneshyari.com/article/392624

Daneshyari.com