Available online at www.sciencedirect.com
SCIENCE DIRECT®
Information Sciences 173 (2005) 227–254
INFORMATION SCIENCES
AN INTERNATIONAL JOURNAL
www.elsevier.com/locate/ins

A segment-wise time warping method for time scaling searching

Mi Zhou *, Man Hon Wong

Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, N.T. Hong Kong, China

Received 2 March 2004

Abstract

The technique of searching for similar patterns among time series data is very important in a wide range of applications. Among them, the time scaling searching is a hard problem that only a few works have tackled. By combining the advantages of a natural time scaling transformation and the dynamic time warping method, we propose a similarity measure that is more suitable for time scaling searching than any existing one. We then explain how to calculate the proposed segment-wise time warping (STW) distance using dynamic programming. In addition, we discuss the lower bound technique of STW distance and the corresponding index method. Through different experiments, we find that the index can greatly reduce the amount of data that must be retrieved, and will lead to great improvements of performance in large sequence database compared with a sequential search.

Keywords: Similarity search; Time warping; Time series database; Spatial indexing

* Corresponding author. Tel.: +852 93183913; fax: +852 26035302.
E-mail addresses: mzhou@cse.cuhk.edu.hk (M. Zhou), mhwong@cse.cuhk.edu.hk (M.H. Wong).

0020-0255/$ - see front matter
doi:10.1016/j.ins.2004.07.014

1. Introduction

Generally, a time series is a sequence of real numbers. The technique of searching for similar patterns among time series data can be applied to many different areas. In most of the applications, the length of the query sequence is the same as that of the result sequences. Formally speaking, given a query sequence $Q(q_1, q_2, \ldots, q_n)$ of length n, the result sequence $S(s_1, s_2, \ldots, s_n)$ should also be of the same length of n. This kind of searching is often called whole matching, and a lot of work has been proposed to handle this problem [6,39,7,17,3,34,2].

However, the query and result sequences may be of different lengths in some applications. This scenario can be further classified into two cases. In one case, although the query and result sequences are of different lengths, the relative difference is very small; we call it near-whole matching. Several papers [37,23,26] have tried to attack this problem based on dynamic time warping.

In the other case, the relative difference is quite big, and sometimes the length of the result sequence can be several or more than ten times greater than that of the query one; we call it time scaling searching. For example, in financial applications, one may request to find all companies whose stock price sequences in a certain period are similar to a given shape such as the so-called *head and shoulders* or *double tops and bottoms* [27,5], in order to predict future trends. Usually, the shape is described by a short sequence while the result sequences can be much longer than the query one. Up to now, only a few papers [25,29,11,23] have addressed this problem.

Moreover, the time scaling searching problem is not just an enumeration of possible cases; it is also important since it is related to self-similarity in time series. Self-similarity [19], which is closely related to power-law [35,19], is a common behavior existing in many artificial and natural phenomena, e.g., financial time series [20], geophysical data [35], and network traffic data [18], etc. Recently, self-similarities in data mining have attracted researchers' attention and will become a fascinating research topic. However, it is not the focus of this paper.

In [25,29], to handle the time scaling searching, the authors in fact rely on a uniform time scaling transformation to make two sequences have the same length, then various similarity measures used in whole matching can be applied. The transformation can be defined as follows: given a sequence $S(s_1, s_2, \ldots, s_n)$ of length n and a scaling factor $a \in N^+$, the transformed sequence S' is of length $a \times n$ such that

$$s'_{a\times(i-1)+1} = s'_{a\times(i-1)+2} = \cdots = s'_{a\times i} = s_i, \quad 1 \leqslant i \leqslant n$$

That is, every point in the original sequence will be uniformly repeated a times to form the scaled sequence; we call it uniform point-wise time scaling. This time scaling transformation does have some nice properties, such as the mean value of the scaled sequence remaining the same as that of the original one.