Temporal Data Mining Using Multilevel-Local Polynomial Models

Weiqiang Lin[1], Mehmet A. Orgun[1], and Graham J. Williams[2]

[1] Department of Computing, Macquarie University Sydney, NSW 2109, Australia, Email: {wlin,mehmet}@comp.mq.edu.au
[2] CSIRO Mathematical and Information Sciences, GPO Box 664, Canberra ACT 2601, Australia, Email: Graham.Williams@cmis.csiro.au

Abstract. This study proposes a data mining framework to discover qualitative and quantitative patterns in discrete-valued time series(DTS). In our method, there are three levels for mining temporal patterns. At the first level, a structural method based on distance measures through polynomial modelling is employed to find pattern structures; the second level performs a value-based search using local polynomial analysis; and then the third level based on multilevel-local polynomial models(MLPMs), finds global patterns from a DTS set. We demonstrate our method on the analysis of "Exchange Rates Patterns" between the U.S. dollar and Australian dollar.

1 Introduction

Discovering both qualitative and quantitative temporal patterns in temporal databases is a challenging task for research in the area of temporal data mining. Although there are various results to date on discovering periodic patterns and similarity patterns in discrete-valued time series (DTS) datasets (e.g. [5, 2, 3]), a general theory and method of data analysis of discovering patterns for DTS data analysis is not well known.

The framework we introduce here is based on a new model of DTS, where the qualitative aspects of the time series are analysed separately to the quantitative aspects. This new approach also allows us to find important characteristics in the DTS relevant to the discovery of temporal patterns. The first step of the framework involves a distance measure function for discovering structural patterns (shapes). In this step, the rough shapes of patterns are only decided from the DTS and a distance measure is employed to compute the nearest neighbors (NN) to, or the closest candidates of, given patterns among the similar ones selected. In the second step, the degree of similarity and periodicity between the extracted patterns are measured based on local polynomial models. The third step of the framework consists of a multilevel-local polynomial model analysis for discovering all temporal patterns based on results of the first two steps which are similarity and periodicity between the structure level and pure value level. We also demonstrate our method on a real-world DTS.

K.S. Leung, L.-W. Chan, and H. Meng (Eds.): IDEAL 2000, LNCS 1983, pp. 180–186, 2000.

The rest of the paper is organised as follows. Section 2 presents the definitions, basic methods and our new method of multilevel-local polynomial models(MLPM). Section 3 applies new models to "Daily Foreign Exchange Rates" data. The last section concludes the paper with a short summary.

2 Definitions and Basic Methods

We first give a definition for what we mean by DTS, and some other notations will be seen later. Then the basic models and our new method which we called multilevel-local polynomial model will be given and studied in detail in the rest of the paper.

2.1 Definitions and Properties

Definition 1 *Suppose that $\{\Omega, \Gamma, \Sigma\}$ is a probability space, and T is a discrete-valued time index set. If for any $t \in T$, there exists a random variable $\xi_t(\omega)$ defined on $\{\Omega, \Gamma, \Sigma\}$, then the family of random variables $\{\xi_t(\omega), t \in T\}$ is called* ***a discrete-valued time series(DTS)****.*

The random variables $\xi_t(\omega)$, $t \in T$ in the above definition should be understood as complex-valued variables in general, and in a sequel, a succinct form of stochastic process is $\{\xi_t(\omega), t \in T\}$, the element ω will be omitted.

In a DTS, we assume that for every successive pair of two time points: t_{i+1} - $t_i = f(t)$ is a function of *time*. For every successive three time points: X_j, X_{j+1} and X_{j+2}, the triple value of (Y_j, Y_{j+1}, Y_{j+2}) has only nine distinct states (or, called nine local features). If we let states: S_s is the same state as prior one, S_u is the go-up state compare with prior one and S_d is the go-down state compare with prior one, then we have state-space $\mathcal{S}$ = {s1, s2, s3, s4, s5, s6, s7, s8, s9} = {(Y_j, S_u, S_u), (Y_j, S_u, S_s), (Y_j, S_u, S_d), (Y_j, S_s, S_u), (Y_j, S_s, S_s), (Y_j, S_s, S_d), (Y_j, S_d, S_u), (Y_j, S_d, S_s), (Y_j, S_d, S_d) }.

Definition 2 *If let $h = \{h_1, h_2, \ldots\}$ be a sequence. If for every $h_j \in h$, $h_j \in \mathcal{S}$, then the sequence h is called a Structural Base sequence. Let $y = \{y_1, y_2, \ldots.\}$ be a real value sequence, then y called a value-point process.*

A sequence is called a *full periodic sequence* if its every point in time contributes (precisely or approximately) to the cyclic behavior of the overall time series (that is, there are cyclic patterns with the same or different periods of repetition). A sequence is called a *partial periodic sequence* if the behavior of the sequence is periodic at some but not all points in the time series.

We have the following results [1]:

Lemma 1. *If let $h = \{h_1, h_2, \ldots\}$ be a sequence and every $h_j \in h$, $h_j \in \mathcal{S}$. If h is a periodic sequence, then h is a structural periodic sequence (existence periodic pattern(s)).*

[1] The proofs are straightforward from above definitions