Temporal Data Mining Using Hidden Markov-Local Polynomial Model

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}@ics.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 similarity and periodicity patterns. At the first level, a structural-based search based on distance measure models is employed to find pattern structures; the second level performs a value-based search on the discovered patterns using local polynomial analysis; and then the third level based on hidden Markov-local polynomial models (HMLPMs), finds global patterns from a DTS set. We demonstrate our method on the analysis of “Exchange Rates Patterns” between the U.S. dollar and the United Kingdom Pound.

Keywords. temporal data mining, discrete-valued time series, similarity patterns, periodicity analysis, local polynomial modelling, hidden Markov models.

1 Introduction

Temporal data mining is concerned with discovering qualitative and quantitative temporal patterns in a temporal database or in a discrete-valued time series (DTS) dataset. DTS commonly occur in temporal databases (e.g., the weekly salary of an employee, or a daily rainfall at a particular location). We identify two kinds of major problems that have been studied in temporal data mining:

1. The similarity problem: finding fully or partially similar patterns in a DTS, and
2. The periodicity problem: finding fully or partially periodic patterns in a DTS.

Although there are various results to date on discovering periodic patterns and similarity patterns DTS datasets (e.g. [4]), a general theory and general method of data analysis of discovering patterns for DTS data analysis is not well known.

Our proposed framework is based on a new model for discovering patterns by using hidden Markov models and local polynomial modelling. The first step of the framework consists of 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 is measured based on local polynomial models. The third step of the framework consists of a hidden Markov-local polynomial model for discovering all levels patterns based on results from the first two steps.

The paper is organised as follows. Section 2 presents the definitions and basic methods of hidden Markov models and local polynomial modelling. Section 3 presents our new method of hidden Markov-local polynomial models (HMLPM). Section 4 applies new models to "Daily Foreign Exchange Rates" data and section 5 discusses related work. The final section concludes the paper with a short summary.

2 Definitions and Basic Methods

We first give a definition of what we mean by DTS and some other notations will be introduced later. The basic models will be given here and studied in detail in the rest of the paper.

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)****.*

2.1 Definitions and Properties

We consider the bivariate data $(X_1, Y_1), \ldots, (X_n, Y_n)$ which form an independent and identically distributed sample from a population (X, Y). For given pairs of data (X_i, Y_i), for $i = 1, 2, \ldots, N$, we can regard the data as generated from the model

$$\mathbf{Y} = m(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon$$

where $E(\varepsilon) = 0$, $Var(\varepsilon) = 1$, and X and ε are independent.

We assume that for every successive pair of two time points in DTS, $t_{i+1} - t_i = f(t)$ is a function (in most cases, $f(t)$ = constant). For every successive three observations: 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 (called local features) depending on changes in value.

Let state: S_s be the same state as the prior one, S_u the go-up state compared with the prior one and S_d the go-down state compared with the prior one, then we have state-space $\mathcal{S}$ = {s1, s2, s3, s4, s5, s6, s7, s8, s9} = $\{(Y_j, S_u, S_u), (Y_j, S_u, S_s), (Y_j, S_u, S_d), (Y_j, S_s, S_u), (Y_j, S_s, S_s), (Y_j, S_s, S_d), (Y_j, S_d, S_u), (Y_j, S_d, S_s), (Y_j, S_d, S_d)\}$.

A sequence is called a *full periodic sequence* if 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.

Definition 2 *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 and a subsequence of h is called a sub-Structural Base sequence. If any subsequence h_{sub} of h is a periodic sequence, then h_{sub} is called a sub-structural periodic sequence, h also is a structural periodic sequence (existence periodic pattern(s)).*

Definition 3 *Let $y = \{y_1, y_2, \ldots.\}$ be a real valued sequence. Then y is called a value-point process. For y_j with $0 \leq y_j < 1$ (mod 1) for all j, we say that y is uniformly distributed if every subinterval of [0, 1] gets its fair share of the terms of the sequence in the long run.*

Definition 4 *Let $y = \{y_1, y_2, \ldots.\}$ be a sequence of real numbers with $I - \delta < y_k < I + \delta$, for all k, where I is a constant and δ is an allowable variable parameter. We say that y has an approximate constant sequence distribution of $y = \{I, I, \ldots.\}$. In general, if $h(t) - \delta < y_k < h(t) + \delta$ for all k, we say that y has an approximate distribution function h(t).*

2.2 Hidden Markov Models (HMMs)

In a hidden Markov model (HMM) an underlying and unobserved sequence of states follows a Markov chain with a finite state space and the probability distribution of the observation at any time is determined only by the current state of that Markov chain. In this subsection we briefly introduce the hidden Markov time series models which is limited to standard results taken from the literature. We have in particular used those of Baldi and Brunak [13].

Let $\{S_t : t \in \mathsf{N}\}$ be an irreducible homogeneous Markov chain on the state space $\{1, 2, \ldots, m\}$, with transition probability matrix Δ. That is, $\Delta = (\eta_{ij})$, where for all states i and j, and times t:

$$\eta_{ij} = \mathsf{P}(\ S_t = j \mid S_{t-1} = i\ )$$

For $\{S_t\}$, there exists a unique, strictly positive, stationary distribution $\gamma = (\gamma_1, \ldots, \gamma_m)$, where we suppose $\{S_t\}$ is stationary, so that γ is, for all t, the distribution of S_t

Suppose there exists a nonnegative random process $\{\xi_t;\ t \in \mathsf{N}\}$ such that, conditional on $S^{(T)} = \{S_t : t = 1, \ldots, T\}$, the random variables $\{\xi_t : t = 1, \ldots, T\}$ are mutually independent and, if $S_t = i$, ξ_t takes the value v with probability π^t_{vi}. That is, for $t = 1, \ldots, T$, the distribution of ξ_t conditional on $S^{(T)}$ is given by

$$\mathsf{P}(\xi_t = v | S_t = i) = \pi^t_{vi}$$

where the probabilities π^t_{vi} as the "state-dependent probabilities". If the probabilities π^t_{vi} do not depend on t, the subscript t will be omitted.

2.3 Local Polynomial Models (LPMs)

The key idea of local modelling is explained in the context of least squares regression models. We use standard results from the local polynomial analysis theory which can be found from the literature on linear polynomial analysis (e.g, [8]). Recall the data model function given earlier: $\mathbf{Y} = m(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon$ where $E(\varepsilon) = 0$, $Var(\varepsilon) = 1$, and X and ε are independent [1]. We approximate the unknown regression function $m(x)$ locally by a

[1] We always denote the conditional variance of Y given $X = x_0$ by $\sigma^2(x_0)$ and the density of X by $f(\cdot)$

polynomial of order p in a neighbourhood of x_0,

$$m(x) \approx m(x_0) + m'(x_0)(x - x_0) + \ldots, + \frac{m^{(p)}(x_0)}{p!}(x - x_0)^p.$$

This polynomial is fitted locally by a weighted least squares regression problem:

$$\text{minimize}\{\sum_{i=1}^{n}\{Y_i - \sum_{j=0}^{p}\beta_j(X_i - x_0)^j\}^2 K_\delta(X_i - x_0)\},$$

where δ is the same δ as in definition 4, and $K_\delta(\cdot)$ with K a kernel function assigning weights to each datum point [2].

3 Hidden Markov-Local Polynomial models (HMLPMs)

A real-world temporal dataset may contain different kinds of patterns such as complete and partial similarity patterns and periodicity patterns, and complete or partial different order patterns. There are many different techniques for efficient sequence or subsequence matching to find patterns in discrete-valued time series database (DTSB) (e.g, [1]). A limitation of those techniques is also that they do not provide a coherent language for expressing prior knowledge and handling uncertainty in the matching process. Also the existence of different patterns does not guarantee the existence of an explicit model.

In this section we introduce our new data mining model for pattern analysis in a DTS by a combination of the hidden Markov models (HMMs) and local polynomial models (LPMs), called hidden Markov-local polynomial models (HMLPMs). HMMs have been successfully used in many applications, such as in isolated word recognition (see [7]), but they have two major limitations. One is HMMs often have a large number of unstructured parameters, and the other is they cannot express dependencies between hidden states. In order to overcome the limitations of HMMs we apply local polynomial modelling techniques to relax the restrictive form of a HMM. We combine HMMs and LPMs to form hybrid models that contain the expressive power of artificial LPMs with the sequential time series aspect of HMMs.

For building up our new data mining model we divide the data sequence or data vector sequence into two groups: (1) the structural-base data group and (2) the pure value-based data group. In group one we only consider the data sequence as a 9-state structural sequence by applying a distance measure function for performing structural pattern search. In group two, we use local polynomial techniques on the pure value-based sequence data for discovering pure value-based patterns. Then we combine those two groups by using hidden Markov models to obtain the final results.

[2] In section 4, we choose Epanechnikov kernel function: $K(z) = \frac{3}{4}(1 - z^2)$ for our experiments in pure-value pattern searching.

3.1 Modelling DTS

Without loss of generality we assume that for each successive pair of time points in a DTS, we have t_{i+1} - t_i = c (a unit constant). According to our method the structural base sequence and value-point process data model become:

$$\mathbf{U} = m(\mathbf{V}) + \sigma(\mathbf{V})\varepsilon$$

where **U** is the number of y_j of a given sample sequence.

Firstly we may view the structural base as a set of vector sequence $\{\mathbf{V_1}, \cdots, \mathbf{V_m}\}$, where each $\mathbf{V_i} = (s1, s2, s3, s4, s5, s6, s7, s8, s9)^T$ denotes the 9-dimensional observation on an object that is to be assigned to a prespecified group.

Then we may also view the value-point process model as a local polynomial model:

$$y(x) = \beta_0 + \beta_1(x - x_0) + \ldots, +\beta_p(x - x_0)^p + \varepsilon.$$

It is more convenient to work with matrix notation for the solution to the above least squares problem in section 2.3. Let

$$\mathbf{X} = \begin{pmatrix} 1 & (X_1 - x_0) & \cdots & (X_n - x_0)^p \\ \vdots & \vdots & & \vdots \\ 1 & (X_1 - x_0) & \cdots & (X_n - x_0)^p \end{pmatrix},$$

and put $\mathbf{Y} = (Y_1, \cdots, Y_n)^T$ and $\hat{\beta} = (\hat{\beta}_0, \cdots, \hat{\beta}_p)^T$.

Further, let **W** be the $n \times n$ diagonal matrix of the weights:

$$\mathbf{W} = diag\{K_\delta(X_i - x_0)\}.$$

The solution vector is provided by weighted least squares theory and is given by

$$\hat{\beta} = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{Y}.$$

Then the problem of value-point pattern discovery can be formulated as the local polynomial analysis of discrete-valued time series.

3.2 Structural Pattern Discovery

We now introduce an approach to discovering patterns in structural base sequences which uses a distance measure function with its density estimator.

From the point of view of our method in structural sequence data analysis, we use squared distance functions which are provided by a class of positive semidefinite quadratic forms. Specifically, if $\mathbf{u} = (u_1, u_2, \cdots, u_9)$ denotes the 9-dimensional observation of each different distance of patterns in a state on an object that is to be assigned to one of the g prespecified groups, then, for measuring the squared distance between $\mathbf{u}$ and the centroid of the ith group, we can consider the function [3]

$$D^2(i) = (\mathbf{u} - \bar{\mathbf{y}})^{'}\mathbf{M}(\mathbf{u} - \bar{\mathbf{y}})$$

where $\mathbf{M}$ is a positive semidefinite matrix to ensure the $D^2(i) \geq 0$.

3.3 Point-Value Pattern Discovery

Here we introduce an enhancement to the local polynomial modelling approach through functional data analysis. On the value-point pattern discovery, given the bivariate data $(X_1, Y_1), \cdots, (X_n, Y_n)$, one can replace the weighted least squares regression function in section 2.3 by

$$\sum_{i=1}^{n} \ell\{Y_i - \sum_{j=0}^{p} \beta_j (X_i - x_0)^j\} K_h(X_i - x_0)$$

where $\ell(\cdot)$ is a loss function. For the purpose of predicting future values we use a special case of the above function with $\ell_\alpha(t) = |t| + (2\alpha - 1)t$ [3].

3.4 Using HMLPMs for Pattern Discovery

For using HMLPMs in pattern discovery we combine the above two kinds of pattern discovery. In structural pattern searching let the structural sequence $\{V_t : t \in \mathsf{N}\}$ be an irreducible homogeneous Markov chain on the state space $\{s1, s2, \ldots, s9\}$, with the transition probability matrix Δ (see section 2.1 for details).

In value-point pattern searching suppose the pure valued data sequence is a non-negative random process $\{C_t;\ t \in \mathsf{N}\}$ such that, conditional on $V^{(T)} = \{V_t : t = 1, \ldots, T\}$, the random variables $\{C_t : t = 1, \ldots, T\}$ are mutually independent and, if $S_t = i$, C_t takes the value v with probability π^t_{vi}. That is, for $t = 1, \ldots, T$, the distribution of C_t conditional on $V^{(T)}$ is given by

$$\mathsf{P}(C_t = v | V_t = i) = \pi^t_{vi}$$

Suppose that if $V_t = i$, ξ_t has a local polynomial distribution with parameters $n_{p,t}$ (a known positive integer) and p_i. That is, the conditional local polynomial distribution of ξ_t has parameters $n_{p,t}$ and $m(t)$, where

$$m(t) = \sum_{i=1}^{m} p_i W_i(t),$$

and $W_i(t)$ is, as before, the indicator of the event $\{V_t = i\}$. Then we have "state-dependent probabilities" for each nine states ($v = 0, 1, \ldots, n_{p,t}$)

The models $\{\xi_t\}$ are defined as hidden Markov-local polynomial models. In this case there are m^2 parameters: m parameters λ_i or p_i, and $m^2 - m$ transition probabilities η_{ij}, e.g. the off-diagonal elements of Δ, to specify the "hidden Markov chain" $\{S_t\}$.

[3] This is often called *quantile regression*.

4 Experimental Results

This section presents selected experimental results. There are three steps of experiments for the investigation of "Daily Foreign Exchange Rates"[4] analysis of "Exchange Rates Patterns" between the U. S. dollar and the U. K. pound. The data consist of daily exchange rate for each business day between 2 January 1971 and 21 June 1999. The time series is plotted in figure 1.

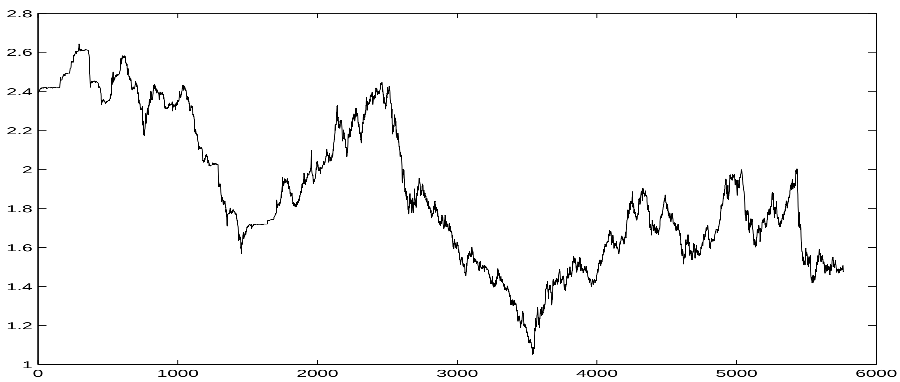

Fig. 1. 5764 working days exchange rates between the U. S. dollar and the U. K. pound, since 1971.

4.1 On structural pattern searching

We investigate the sample of the structural base to test naturalness of the similarity and periodicity on the Structural Base distribution. The size of this discrete-valued time series is about 5764 points. We consider 9 states in the state-space of structural distribution: $\mathcal{S} = \{s1, s2, s3, s4, s5, s6, s7, s8, s9\}$.

In Figure 2, each point represents the occurence of one of the nine transition states, retaining the original order of the states. There exist two approximation uniformly distributed on state 3 and state 7 if the observations are big enough. Figure 2 also explains two facts: (1) there exists a hidden periodic distribution which corresponds to patterns on the same line with different distances, and (2) there exist partial periodic patterns on and between the same lines. To explain this further, we can look at the plot of distances between the patterns at a finer granularity over a selected portion of the daily exchange rates. For instance, in the right of Figure 2 the dataset consists of daily exchange rates for 300 business days starting from 3 January 1983, telling us there exist a number of partial periodic patterns appearing in each year and, also telling us in each state in a

[4] The Federal Reserve Bank of New York for trade weighted value of the dollar = index of weighted average exchange value of U.S. dollar against the United Kingdom Pound: `http://www.frbchi.org/econinfo/finance/finance.html`.

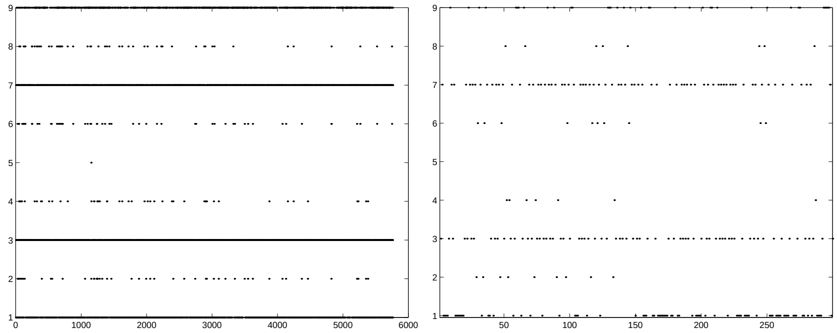

Fig. 2. Left: plot of the distance between same state for all 9 states in 5764 business days. Right: plot of the distance between same state for all 9 states in first 300 business days.

year there is a hidden periodic and similarity distribution with each point representing the distance of patterns of various forms. Between some combined pattern classes there exist similar patterns such as between 5 to 10 and 15 to 22; between 32 to 35 and 42 to 44.

In Figure 3 the *x*-axis represents how many times the same distance is found between repeating patterns and the *y*-axis represents the distance between the first and second occurences of each repeating pattern. In other words, we classify repeating patterns based on a distance classification technique. Again we can look at the plot over a selected portion to observe the distribution of distances in more detail. For example, in the right of figure 3 the dataset consists of daily exchange rates for the first 50 business days. It can be observed that the distribution of distances is a cubic curve distribution: $y = \frac{1}{ax^2+bx+c}$, where Δ = ax^2 + bx + c and Δ, b < 0, a > 0.

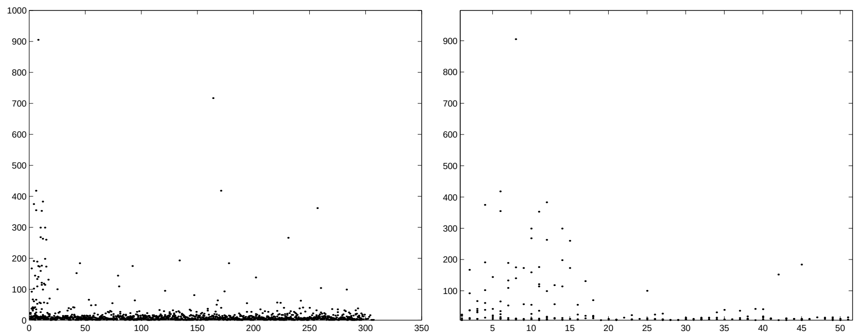

Fig. 3. Left: plot of the distance between same state for all states in 5764 business days. Right: plot different pattern appear in different distances for first 50 business days.

In summary, some results for the structural base experiments are as follows:

- Structural distribution is a hidden periodic distribution with a periodic length function $f(t)$ (there are techniques available to approximate to the form of this function such as higher-order polynomial functions).
- There exist some partial periodic patterns based on a distance shifting.
- For all kinds of distance functions there exist a cubic curve: $y = \frac{1}{ax^2+bx+c}$, where Δ = ax^2 + bx + c and Δ, b < 0, a > 0.
- there exists an approximate uniform distribution in state 3 and state 7.

4.2 On value-point pattern searching

We now illustrate our new method to construct predictive intervals on the value-point sequence for searching periodic and similarity patterns. The linear regression of value-point of X_t against X_{t-1} explains about 99% of the variability of the data sequence, but it does not help us much in analysis and predicting future exchange rates. In the light of our structural base experiments, we have found that the series $Y_t = X_t - X_{t-2}$ has non-trivial autocorrelation. The correlation between Y_t and Y_{t-1} is 0.5268. Then the observations can be modelled as a polynomial regression function, say

$$Y_t = X_t - X_{t-2} + \sigma(X_t)\varepsilon_t, \qquad t = 1, 2, \ldots, N$$

and then the following new series

$$y(t) = Y(t) + Y(t-1) + \varepsilon_{t'} \qquad t = 1, 2, \ldots, N$$

may be obtained. We also consider the $\varepsilon(t)$ as an auto-regression $AR(2)$ model

$$\varepsilon_{t'} = a\varepsilon_{t'-1} + b\varepsilon_{t'-2} + e_{t'}$$

where a, b are constants dependent on sample dataset, and $e_{t'}$ with a small variance constant which can be used to improve the predictive equation. Our analysis is focused on the series Y_t which is presented in the left of Figure 4. It is scatter plot of lag 2 differences: Y_t against Y_{t-1}.

We obtain the exchange rates model according to nonparametric quantile regression theory:

$$Y_t = 0.488Y_{t-1} + \varepsilon_t$$

From the distribution of ε_t, the $\varepsilon(t)$ can be modelled as an $AR(2)$

$$\varepsilon_t = 0.261\varepsilon_{t-1} - 0.386\varepsilon_{t-2} + e_t$$

with a small Var(e_t)(about 0.00093) to improve the predictive equation.

For prediction of future exchange rates for the next 210 business days, we use the simple equation $Y_t = 0.488Y_{t-1}$ with an average error of 0.00135. In the right of Figure 4 the actually observed series and predicted series are shown.

Some results for the value-point of experiments are as follows:

- There does not exist any full periodic pattern, but there exist some partial periodic patterns based on a distance shifting.
- There exist some similarity patterns with a small distance shifting.

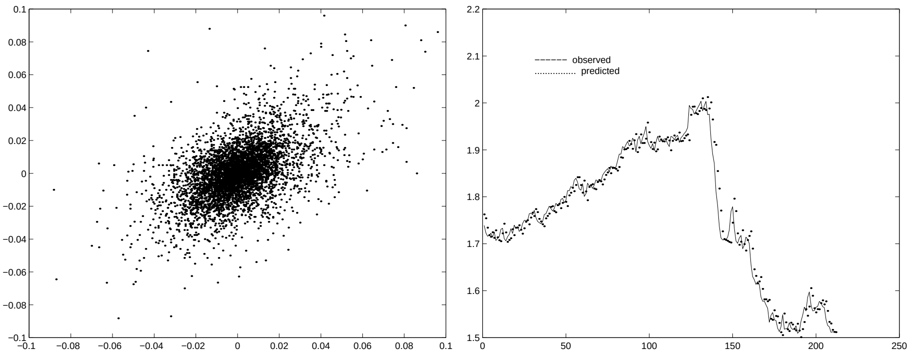

Fig. 4. Left: Scatter plot of lag 2 differences: Y_t against Y_{t-1}. Right:Plot of future exchange rates only for 210 business days by using the simple equation Y_t = 0.488 Y_{t-1}

4.3 Using HMLPMs for pattern searching

Let $\{S_t : S_t \in \mathcal{S}, t \in \mathsf{N}\}$ be an irreducible homogeneous Markov chain on the state space $\{s1, s2, s3, s4, s5, s6, s7, s8, s9\}$, with transition probability matrix (TPM) (or, stochastic matrix) Δ:

$$\Delta = \left| \begin{array}{ccccccccc} 0.5186 & 0.0208 & 0.4606 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5161 & 0.0323 & 0.4516 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.5146 & 0.0362 & 0.4492 \\ 0.4118 & 0.0588 & 0.5294 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0.5 \\ 0.5355 & 0.0260 & 0.4385 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.4359 & 0 & 0.5641 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.4962 & 0.0342 & 0.4696 \end{array} \right|$$

We are interested in the future of distribution of TPM, $f(t) = \Delta^t$. According to the Markov property, the TPM: $\lim_{t\to\infty} \Delta^t = 0$. This means that the TPM is non-recurrent of a state si to a state sj. In other words, we cannot use present exchange rate to predict future exchange rate of some period after, but we are only able to predict near future exchange rate.

Suppose that our prediction of future exchange rate of value-point sequence is a nonnegative random process $\{C_t;\ t \in \mathsf{N}\}$, and satisfy $Z_t = \alpha Z_{t-1} + \theta_t$.

Suppose the distribution of sequence of transition probability matrix (TPM) under time order $\Delta_1, \Delta_2, \cdots, \Delta_t; t \in \mathsf{N}$ corresponding to the prediction value-point $Z_t = Y_t - Y_{t-2}$.

We have main combined-results on exchange rates as follows:

– We are only able to predict a short future period by using all present information.
– There does not exist any full periodic pattern but there exist some partial periodic patterns.

– There exist some similarity patterns with a small distance shifting.

5 Related Work

According to pattern theory objectives in pattern searching can be classified into three categories:

– Create a representation in terms of algebraic systems with probabilistic superstructures intended for the representation and understanding of patterns in nature and science.
– Analyse the regular structures from the perspectives of mathematical theory.
– Apply regular structures to particular applications and implement the structures by algorithms and code.

In recent years various studies have considered temporal datasets for searching different kinds of and/or different levels of patterns. These studies have only covered one or two of the above categories. For example, many researchers use statistical techniques such as Metric-distance based techniques, Model-based techniques, or a combination of techniques (e.g, [14]) to search for different pattern problems such as in periodic patterns searching (e.g., [9]) and similarity pattern searching (e.g., [6]).

Some studies have covered the above three categories for searching patterns in data mining. For instance [2] presents a "shape definition language", called $\mathcal{SDL}$, for retrieving objects based on shapes contained in the histories associated with these objects. Also [12] present a logic algorithm for finding and representing hidden patterns. In [5], authors described adaptive methods which are based on similar methods for finding rules and discovering local patterns.

Our work is different from these works. First, we use a statistical language to perform all the search work. Second, we divide the data sequence or, data vector sequence, into two groups: one is the structural base group and the other is the pure value based group. In group one our techniques are similar to Agrawal's work but we only consider three state changes (i.e., up (value increases), down (value decreases) and same (no change)) whereas Agarwal considers eight state changes (i.e., up (slightly increasing value), Up (highly increasing value), down (slightly increasing value) and so on). In this group, we also use distance measuring functions on structural based sequences which is similar to [12]. In group two we apply statistical techniques such as local polynomial modelling to deal with pure data which is similar to [5]. Finally, our work combines significant information of two groups to get global information which is behind the dataset.

6 Concluding Remarks

This paper has presented a new approach combining hidden Markov models and local polynomial analysis to form new models of application of data mining. The rough decision for pattern discovery comes from the structural level that is a collection of certain predefined similarity patterns. The clusters of similarity patterns are computed in this level by the choice of certain distance measures. The point-value patterns are decided in the second level and the similarity and periodicity of a DTS are extracted. In the final level we combine structural and value-point pattern searching into the HMLPM model to obtain a global pattern picture and understand the patterns in a dataset better. Another approach to find similar and periodic patterns has been reported else where [10, 11]. With these the model used is based on hidden periodicity analysis and plocal polynormial analysis. However, we have found that using different models at different levels produces better results.

The "Daily Foreign Exchanges Rates" data was used to find the similar patterns and periodicities. The existence of similarity and partially periodic patterns are observed even though there is no clear full periodicity in this analysis.

The method guarantees finding different patterns if they exist with structural and valued probability distribution of a real-dataset. The results of preliminary experiments are promising and we are currently applying the method to large realistic data sets such as two kinds of diabetes dataset.

References

1. Rakesh Agrawal, King-Ip Lin, Harpreet S. Sawhney, and Kyuseok Shim. Fast similarity search in the presence of noise, scaling, and translation in time-series databases. In *Proceedings of the 21st VLDB Conference*, Zürich, Switzerland, 1995.
2. Rakesh Agrawal, Giuseppe Psaila, Edward L.Wimmers, and Mohamed Zait. Querying shapes of histories. In *Proceedings of the 21st VLDB Conference*, 1995.
3. T. W. Anderson. *An introduction to Multivariate Statistical Analysis*. Wiley, New York, 1984.
4. C. Bettini. Mining temportal relationships with multiple granularities in time sequences. *IEEE Transactions on Data & Knowledge Engineering*, 1998.
5. H. Mannila G.Renganathan G. Das, K. Lin and P. Smyth. Rule discovery from time series. In *Proceedings of the international conference on KDD and Data Mining(KDD-98)*, 1998.
6. D.Gunopulos G.Das and H. Mannila. Finding similar time seies. In *Principles of Knowledge Discovery and Data Mining '97*, 1997.
7. P. Guttort. *Stochastic Modeling of Scientific Data*. Chapman & Hall, London, 1995.
8. J.Fan and I.Gijbels, editors. *Local polynomial Modelling and Its Applications*. Chapman and hall, 1996.
9. Cen Li and Gautam Biswas. Temporal pattern generation using hidden markov model based unsuperised classifcation. In *Proc. of IDA-99*, pages 245–256, 1999.
10. Wei Q. Lin and Mehmet A.Orgun. Temporal data mining using hidden periodicity analysis. In *Proceedings of ISMIS2000*, University of North Carolina, USA, 2000.
11. Wei Q. Lin, Mehmet A.Orgun, and Graham Williams. Temporal data mining using multilevel-local polynomial models. In *Proceedings of IDEAL2000*, The Chinese University of Hongkong, Hong Kong, 2000.
12. S.Jajodia andS.Sripada O.Etzion, editor. *Temporal databases: Research and Practice*. Springer-Verlag,LNCS1399, 1998.
13. P.Baldi and S. Brunak. *Bioinformatics & The Machine Learning Approach*. The MIT Press, 1999.
14. Z.Huang. Clustering large data set with mixed numeric and categorical values. In *1st Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 1997.