APPLIED HIDDEN PERIODICITY ANALYSIS FOR MINING DISCRETE-VALUED TIME SERIES DATABASES

WEIQIANG LIN AND MEHMET A. ORGUN

Department of Computing, Macquarie University, NSW 2109, Australia
E-mail: {ulin,mehmet}@comp.mq.edu.au

Discovering qualitative and quantitative temporal patterns in temporal databases (that is, temporal data mining) is a challenging area of research. Temporal data mining involves time series analysis on the information held in a temporal database. Discrete-valued time series (DTS) that are common in practice, are sequences of observations on a particular set of variables over time. In this paper, we discuss few major models for DTS that have been proposed in diverse application areas. Then we present a new application of hidden periodicity analysis on a real-world temporal dataset for discovering similar and periodic patterns.

1 Introduction

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

1. The similarity problem: finding fully or partially similar patterns in a DTS or finding fully or partially similar patterns to a given sequence (or query) in a DTS, and

2. The periodicity problem: finding fully or partially periodic patterns in a given DTS.

Although there are many results to date on discovering periodic patterns and similarity patterns in time-series datasets \(^1,2,3,4\), a general theory of discovering patterns for DTS data analysis is not yet well known. Some progress has been made in the temporal database area \(^5,6\). Since there are many different types of discovery problems that have been addressed in the literature, it is important to characterize these problems using some formal framework. Many researchers of temporal data mining pay little or no attention to basic methods and models which are fundamental in discrete-valued time series models. Although there are two popular models, namely the similarity model and periodic model, that have been studied in the past few years, a general
method of analysis is still lacking even for using different measures on a given time series data-set or its samples.

In this paper, we propose a general framework for analysing a DTS and then we focus on the special problem of discovering patterns using hidden periodicity analysis. The framework is based on a new model of DTS that addresses the problem of using a small subset of a real DTS to represent the whole DTSB and to find out important characteristics in the DTS that may lead to the discovery of similar or periodic patterns. We in particular propose a new method for temporal data mining, based on hidden periodicity analysis in noisy data, which combines hypothesis testing method of Grenander and large sample theory.

The rest of the paper is organized as follows. Section 2 briefly reviews some examples in discrete-valued time series and surveys few models that have been proposed. Section 3 briefly introduces the theory and methods of hidden periodicity analysis. Section 4 moves onto experimental analysis for pattern discovery and the paper ends with a brief discussion.

2 Discrete-valued Time Series Examples and Its Models

One of the basic tasks of temporal data mining and temporal data analysis is to automatically discover qualitative and quantitative patterns in a temporal database. According to temporal characteristics, objects in temporal databases can be classified into three categories:

- Time-invariant objects,
- Time-varying objects, and
- Time-series objects.

In rest of the section, we focus only on temporal databases for discrete-valued time-series objects, which are often called time-related databases. Time-related databases are of growing importance in many modern database applications, such as data mining, data warehousing and so on.

2.1 Some Examples in Discrete-valued Time Series

Most of the temporal databases naturally arise in different business organizations as well as scientific decision-support applications. Many of the time series which occur in practice are by their very nature discrete-valued, although it is often quite adequate, and obviously very convenient, to represent them by means of models based on their distributions. For example, most
discrete-valued time series data sets can be represented by means of models based on normal distribution. Some examples of discrete-valued series are:

- stock exchange daily records
- the sequence of wet and dry days at some site
- road accident or traffic counts
- base sequences in DNA
- multiple spectral astronomical data, etc.

Although in some cases models based on a well-defined distribution will suffice (e.g., Poisson distribution), this will not always be so. There are a number of great challenges in mining discrete-valued time series databases. For example, when a time series temporal database is given, discovering inherent patterns and regularities is a challenging task. Another challenge is when no time-domain knowledge is available or its knowledge is too weak; this problem is often called the time-dimension problem 9.

2.2 Discrete-valued Series Models

There are many discrete-valued data models, proposed in applied statistics literature, but these models have not found their way into major data mining research areas. It may be that there is no well-known family of models that are structurally simple, sufficiently versatile to cater for a useful variety of data types, and readily accessible to the practitioner. We briefly outline some discrete-valued time series models which we consider can be developed further and easier to apply in discrete-valued data mining or temporal data mining:

1. Markov model: it uses a simple property of Markov chains on a space of $m$ finite states as possible models for time series taking values in the space, which have $m^2 - m$ independently determined transition probabilities.

2. Hidden Markov model: it uses an unobserved stationary Markov chain and either a Poisson or a binormal distribution.

3. Geometric model: it uses dependence marginal distributions and correlation structure to compare with individual distribution.

4. State-space model: it uses a distribution of observations with mean given by a distribution process.

Although there are two kinds important problems such as similarity and periodic problems that have been studied in recent years, yet general standard
methods are not there even for using a different measure on data-set or its samples. We propose a new model of DTS that addresses the problem using a small sample set to produce the whole temporal data set. The model is based on hidden periodicity analysis for discovering temporal patterns in temporal databases and related to the problem of similarity via multiple models.

3 Using Hidden Periodicity Analysis in DTS

This section introduces hidden periodicity analysis in DTS which will be studied in detail in the rest of the paper. First of all, we give a definition for what we mean by a discrete-valued time series; some other notations will be given in the sequel.

3.1 Definition and Methods

We now introduce some basic ideas and models of our new method. Those ideas and models can be used for mining (or discovering) all different kinds of patterns in both multivariate discrete-valued time series and multidimensional discrete-valued time series.

Definition 1 Suppose that \( \{\Omega, \Gamma, \Sigma\} \) is a probability space, where \( \Omega \) is a set, called the probability, or the sample space; \( \Gamma \) is its subsets, called the events; \( \Sigma \) is probability measure on \( \{\Omega, \Gamma\} \) and \( T \) is an 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 \( \{\xi_t(\omega), t \in T\} \) will be called a stochastic process. If \( T \) is a discrete-valued time index set, then the family of random variables \( \{\xi_t(\omega), t \in T\} \) will be called a discrete-valued time series.

Consider the bivariate data \( (X_1, Y_1), \ldots, (X_n, Y_n) \), which form an independent and identically distributed sample from a population \((X, Y)\). Of interest is to estimate the regression function \( m(x_0) = E(Y|X = x_0) \) and its derivatives. For given pairs of data \((X_i, Y_i), i = 1, 2, \ldots, N\), we can regard the data as being generated from the model

\[
Y = m(X) + \sigma(X)\varepsilon
\]

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

Without loss of generality, we assume that for every successive two time points in DTS \( t_{i+1} - t_i = f(t) \) is a function (e.g., in most cases \( f(t) = \text{constant} \)). For every successive three time points: \( X_j, X_{j+1} \) and \( X_{j+2} \), the triple \((Y_j, Y_{j+1}, Y_{j+2})\) has only 9 distinct states (or local features). Given a value, we identify three possible changes in value: \( S_1 \) is the same state as the prior one (i.e., no change in value), \( S_2 \) is the go-up state compared with the prior one
(i.e., the value increases) and $S_3$ is the go-down state compared with the prior one (i.e., the value decreases). Then we have the state-space $S = \{s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9\} = \{(Y_j, S_2, S_2), (Y_j, S_2, S_1), (Y_j, S_2, S_3), (Y_j, S_1, S_2), (Y_j, S_1, S_1), (Y_j, S_1, S_3), (Y_j, S_3, S_2), (Y_j, S_3, S_1), (Y_j, S_3, S_3)\}$.

**Definition 2** Let $w = \{w_0, w_1, \ldots\}$ be a sequence. If a subset of $w$, say $w_{k+1} = \{w_{k+1}, \ldots, w_{k+1}\}$ ($i \geq 0$ and an integer) contributes to a function $h(j)$ ($j \in 0, 1, \ldots$), then $w$ is called a periodic sequence (or periodic pattern) with periodicity function $h(j)$.

**Definition 3** Let $B = \{S_1, S_2, \ldots\}$ be a sequence. Let $B_{sub} = \{S_1, S_2, \ldots, S_N\}$, where $S_j \in S, j = 1, 2, \ldots, N$ and $N$ be the size of the subset. Then $B_{sub}$ is called the sub-Structural Base of a sample. Also it may be called the sub-Structural Base of $B$ under the same direction of sequence.

If $B_{sub} = \{S_1, S_2, \ldots, S_N\}$ is a periodic sequence, then $B_{sub}$ may called a substructural periodic sequence (e.g., full or partial structural periodic pattern). Also $B$ is a structural periodic sequence.

**Definition 4** Let $Y = \{y_1, y_2, \ldots\}$ be a real valued sequence. If $y_{sub} = \{y_1, y_2, \ldots, y_N\}$ ($y_j \in Y, j = 1, 2, \ldots, N$ and $N$ is the size of a subset of $Y$), then it may be called a value-point process.

If $y_j$ with $0 \leq y_j < 1$ (mod 1) for all $N$. 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. More precisely, if

$$\lim_{n \to \infty} \frac{\text{number of } \{j \leq n : y_k \in J\}}{n} = \text{length of } J$$

for all subintervals $J$ of $[0, 1)$, or in the point of statistical probability,

$$P(\xi \in J) = \frac{\nu_d(J)}{\nu_d(Y)}$$

for all Borel sets $J$ contained in $Y$.

A uniformly distributed discrete-valued time series is a rather trivial random pattern. However, $M$ independent uniformly distributed datasets can be superposed to form a new dataset pattern. In the general case, independent statistical distribution $Y_j$ of a dataset in combination with Structural Base distribution $B$ of the dataset can be superposed to form every different kinds of patterns in the dataset.

**Definition 5** 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 an integer, $\delta > 0$ and $I \gg \delta$). 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 $h(t)$.
We have the following result:

**Lemma 1** A discrete-valued dataset contains periodic patterns if and only if there exist structural periodic patterns and periodic value-point processes with or without an independently identical distribution (i.i.d.).

When looking for periodic patterns, we use classification techniques both in value-point process and base structural process. In recent years, a lot of work have been done in classification areas by using “Statistical Approaches to Predictive Modeling”. Our new method also includes estimating a categorical variable 10:

1. Density Estimation, such as kernel density estimators 11.
2. Metric-space based methods, such as K-nearest-neighbor method 11, and
3. Projection into decision regions.

If there is a large number of groups of variables, then we use multidimensional classification:

- the choice of an effective space or representation for discrimination, and
- the choice of a distance measure or metric for use in such a space.

For example, given a large number of groups $g$, each of which has $p$-dimensional space, an “unknown” is classified by comparing its distances from all of the group centroids.

**Lemma 2** In a discrete-valued dataset, there exist similarity patterns if and only if there exist structural base periodic patterns and similarity value-point distribution with or without an independently identical distribution.

When looking for similarity patterns, our new method uses similarity and optimization of value-point and base structural set functions. It includes the cases when the number of clusters is given or not given.

If the number of clusters is given, then we may use the following techniques:

- Metric-distance based technique: a distance measure is defined and the objective becomes finding the best $k$-way partition such that each block of the partition is closer to each other (or centroid) than to cases in other partition.
- Model-based technique: a model is hypothesized for each of the clusters and the idea is to find the best fit of that model to each cluster (using techniques such as maximum likelihood, linear or non-linear regression with or without parameters, etc).
• Partition-based technique: it basically enumerates various partitions and then scores them by some criterion.

The techniques above can be occasionally used in combination, such as Probability-based vs. Distance-based can be combined to Probability Distance-based clustering analysis.

If the number of clusters is unknown, then we can use Non-Hierarchical Clustering Algorithms to find their $k^{12,13}$.

3.2 Theory of Hidden Periodicity Analysis

We first introduce the hypothesis testing method of Grenander $^7$ for detecting hidden periodicities in noisy data. Suppose that the model of general observations of a sub-set is that of

$$x(t) = \sum_{n=1}^{P} \xi_n e^{i\lambda_n t} + \eta(t), \quad t \in \mathbb{Z}$$

where $P$ is known, $\xi_n$ and $\lambda_n$ are unknown parameters, $\eta(t)$ is independent identically distributed (i.i.d). $\mathcal{N}(0, \sigma^2)$ and $\sigma$ is unknown parameter. But Grenander considered the model

$$\xi(t) = \sum_{k=1}^{P} A_k \cos(\omega_k t) + \gamma(t)$$

where $P$ is known, $A_k$, $\omega_k$ are unknown parameters, $\gamma(t)$ is i.i.d. $\mathcal{N}(0, \sigma^2)$ and $\sigma$ is unknown parameter. Let sub-set size $M = 2m + 1$ and put

$$\hat{I}_M = \frac{2}{M} \left| \sum_{k=1}^{M} \xi(k) e^{-ik\lambda} \right|^2$$

$$\hat{I}_p = \hat{I}_M \left( \frac{2p\pi}{N} \right), \quad 1 \leq p \leq m.$$ 

Then Grenander $^7$ suggested the following testing procedure:

$$H_0 : \xi(t) = \gamma(t), \quad t = 0, \pm 1, \pm 2, \ldots.$$ 

$$H_1 : M = r (\xi(t) possesses r frequency components), 1 \leq r \leq m.$$ 

Let $\hat{I}(r)$ be the $r$th largest value of $\{\hat{I}_r\}$, put

$$g(r) = \frac{I(r)}{\sum_{p=1}^{m} \hat{I}_p}$$
then the distribution of the statistic \( g(r) \) is

\[
P\{ g(r) > z \} = \frac{m!}{(r-1)!} \sum_{j=r}^{\lfloor \frac{r}{2} \rfloor} \frac{(-1)^{j-r}(1-jz)^{M-1}}{j(m-j)!(j-r)!}.
\]

If the test rejects the \( H_0 \), then it shows that \( \xi(t) \) is not a pure white noise series and possesses \( r \)-frequency components. Suppose that the estimates \( \{\hat{I}(p)\} \) have been arranged in the order

\[
\hat{I}(1) \geq \hat{I}(2) \geq \ldots \geq \hat{I}(r)
\]

and the corresponding frequencies are

\[
\hat{\lambda}_k = k \frac{2\pi}{N}, \quad k = 1, 2, \ldots, r
\]

then \( \{\hat{\lambda}_k, k = 1, 2, \ldots, r\} \) are estimates of hidden frequencies in \( \xi(t) \), and \( r \) is the order.

The amplitudes \( \hat{A}_k \) vs. \( \hat{\lambda}_k \) can be estimated by

\[
\hat{A}_k = \frac{1}{N} \sum_{n=1}^{N} \xi(n) e^{in\hat{\lambda}_k}, \quad k = 1, 2, \ldots, r
\]

Note: In the previous hypothesis testing, the parameter \( r \) is assumed to be known a priori. Since in the usual cases \( r \) is unknown, we use the testing procedure step by step, i.e., first put \( r = 1 \) and do the testing. If \( H \) is rejected then we put \( r = 2 \) and so on, until, say \( r = p + 1 \), when \( H \) is accepted then we estimate the order as \( p \). If \( r = 1 \) is rejected, then it means \( \xi(t) \) is not a white noise series, the distribution of \( P\{ g(r) > z \} \) has to be changed.

4 Pattern Discovery

A real-world temporal dataset may contain different kinds of patterns such as complete or partial similarity pattern and periodicity pattern, complete or partial different order patterns. There are many different techniques for efficient sequence or subsequence matching to find patterns in DTSB \(^2\). But there is a limitation of those techniques in general is that they do not provide a coherent language for expressing prior knowledge and handling uncertainty in the matching process \(^4\). Also the existence of a different pattern does not guarantee the existence of an explicit form. In this section, we present our new method of pattern discovery on a DTSB by using Hidden periodicity analysis.
4.1 Modeling DTS

One of the aims in time series analysis is to find a statistical model for a given data-set. Without loss of generality, we assume that for every successive two time points in DTS have $t_{i+1} - t_i = c$ (a unit constant). Recall the data generated model:

$$Y = m(X) + \sigma(X)\varepsilon$$

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

According to our new method, we may assume the value-point process data model is a linear model

$$Y = X\beta + \varepsilon$$

The linear model based upon least square estimation (LSE) is

$$Y' = X'\beta + \varepsilon'$$

$$\hat{\beta} = (X^TX)^{-1}X^TY$$

According to linear model theory, we have

$$\hat{\beta} \sim N(\beta, Cov(\hat{\beta}))$$

Particularly, for $\hat{\beta}_i$ we have

$$\hat{\beta}_i \sim N(\beta_i, \sigma_i^2),$$

where $\sigma_i^2 = \sigma^2 a_{ii}$, and $a_{ii}$ is the $i-th$ diagonal element of $(X^TX)^{-1}$.

Now, for each value-point, we may fit a linear model as above and parameters can be estimated under LSE. Since $\hat{\beta}_i$ follows normal distributions, we may have the $t$-test for the means of each given sample set (or two subsets). If we cannot reject the hypothesis, then we have possibility that the features remain uncovered in the fluctuation periods of observation curves. Therefore, we first remove the trend effect of each curve from the original record by subtracting the above regression function at $x$ from the corresponding value to obtain a comparatively stationary series.

Then the problem can be formulated as the hidden periodicity analysis of discrete-valued time series. Suppose that the model of observations is that of

$$x(t) = \sum_{n=1}^{P} \xi_n e^{i\lambda_n t} + \eta(t), \quad t \in Z$$

where $\{\lambda_n\}$ are bounded random variables, $P$ is an unknown parameter and $\eta(t)$ is a weak-$P$ model so that the correlated stationary series may be included.
Note: According definition of "Weak - P series", the following series are $Weak - P$ series:

- $\eta(t)$ is a Gaussian stationary series
- $\eta(t)$ is i.i.d. series with a fourth order moment
- $\eta(t)$ is a linear process

For a sub-DTS say, $x_j(t), (j=1, 2, \ldots, N)$ we first transform it into the frequency domain by applying the discrete-valued Fourier Transform (DFT) and also we have the spectral power of the sub-DTS for every frequency bin.

4.2 Experimental Results

Due to space limitations we only present a small subset of our experimental results. There are two groups of experiments for the investigation of "Daily Foreign Exchange Rates" \(^a\) analysis of "Exchange Rates Patterns" between U.S. dollar and Canadian dollar (the exchange data is plotted in figure 1):

- structural pattern searching, and
- exchange rate pattern searching.

The main results are:

On structural pattern searching:

We are investigating the sample of structural base (eg, one year based) to test the naturalness of the similarity and periodicity on Structural Base distribution. The accumulation of each state every year satisfies a similar linear function (see Figure 2):

$$S(t) = \beta_0 + \beta_1 t$$

We then adjust off the trend by the linear regression function $S(t)$ and obtain a new series

$$z(t) = x(t) - S(t), \quad t = 1, 2, \ldots, N$$

where $x(t)$ is the original record.

Then we may use the hidden periodicity analysis for the new structural base series $z(t)$. The algorithm is hybrid, i.e. it is based on the asymptotic theory of large samples of time series then connected to other algorithms.

\(^a\)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 Canada dollar: http://www.frbchi.org/econinfo/finance/finance.html
Figure 1. 11 years of daily U.S. dollar exchange rate against Canadian dollar, originally 2735 points since 1971.

Figure 2. The accumulation of states over 11 years of daily U.S. dollar against Canadian dollar in structural base, originally 2735 points since 1971

Preliminary results have shown us that:

- The structural base series for each state in a year has a hidden periodic distribution. For instance, the number of states each year in structural base is a linear and approximate uniform distribution (see figure 2).
there exist a number of partial periodic patterns appearing in each year, also it may have a logarithmic normal distribution (figures 3 and 4).

For instance in Figure 4, the dataset consists of daily exchange rates for 200 business days starting from 3 January 1972. Each point represents the occurrence of one of nine transition states, retaining the original order of the states. Between states 3 and 7 there exist similar patterns such as between 10 to 12 and 132 to 134, etc. It can also be observed that each state (e.g, state 3) when considered in conjunction with another state (e.g, state 7) forms a hidden periodic distribution.

**On exchange rate pattern searching:**

The data consist of daily exchange rates at every business day between U.S. dollar versus Canadian dollar. The length of this time series is about 7037 (1971 - 1998) and following analysis is carried out on the observed data.

Some results for the value-point of experiments are given below:

- value-point distribution is a non-linear distribution, but it is a hidden uniformly distributed with very small error,
- there does not exist any full periodic pattern, but there exist some partial periodic patterns based on distance shifting,
Figure 4. 200 business days of daily U.S. dollar exchange rate against Canadian dollar in states base

- there exist similarity patterns, etc.

For instance, consider the 3 years of daily U.S. dollar exchange rate against Canadian dollar (1978-1980). Suppose the exchange rate value can be modeled as

\[ Y_i = m(t)Y_{i+k} + \varepsilon_i, \quad (k > 0 \text{ and fixed integer}) \]

We then adjust off the trend by the linear regression function \( Y_i \) and a new series

\[ w(t) = v(t) - Y(t), \quad t = 1, 2, \ldots, N \]

may be obtained, where \( v(t) \) is the original record. The entire dataset is plotted in Figure 5.

And we consider the \( w(t) \) as a weak-\( P \) linear model

\[ w(t) = \sum_{n=1}^{P} A_n \cos \left( \frac{2\pi}{Y_n} t + \theta_n \right) + \eta(t), \]

where

\[ P, \{ S_n, \theta_n, A_n, n = 1, 2, \ldots, P \} \]

are unknown parameters, \( \eta(t) \) is a stationary correlated residual series satisfying the condition of weak-\( P \).
Figure 5. Adjusting off the trend of each year by a linear function in 11 years of daily U.S. dollar exchange rate against Canadian dollar in value-point.

Then we may use the hidden periodicity analysis for the new structural base series \( w(t) \). The algorithm is hybrid, i.e. it is based on the asymptotic theory of large samples of time series then connected to other algorithms.

5 Conclusion

This paper has presented a new method for temporal data mining based on Hidden Periodicity Analysis for finding patterns in discrete-valued time series database. The work described in this paper is still in its preliminary stages. In the near future, we plan to extend the method to other kinds of statistical models such as hidden Markov models, polynomial regression models and so on. Our method guarantees finding different patterns with structural and valued probability distribution of a real-dataset. It can be implemented using a straightforward algorithm programming idea, the results of preliminary experiments are very promising.

Acknowledgments

This research has been supported in part by an Australian Research Council (ARC) grant and a Macquarie University Research Grant (MURG). Thanks are also due to Chit Swe and Graham Williams for many fruitful discussions.
References


