Mining Temporal Patterns from Health Care Data

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

[1] Department of Computing, I.C.S., 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@csiro.au

Abstract: This paper describes temporal data mining techniques for extracting information from temporal health records consisting of a time series of elderly diabetic patients' tests. Diabetes is one of the most common diseases affecting quality of life and is potentially life threatening for elderly people in developed countries. We propose a data mining procedure to analyse these time sequences in three steps to identify patterns from any longitudinal data set. The first step is a structural-based search using wavelets to find pattern structures. The second step employs a value-based search over the discovered patterns using the statistical distribution of data values. The third step combines the results from the first two steps to form a hybrid model. The hybrid model has the expressive power of both wavelet analysis and the statistical distribution of the values. Global patterns are therefore identified.

Keywords: temporal data mining, discrete-valued time series, similar patterns, periodicity analysis, waveletes function, data statistical distribution.

1 Introduction

Temporal data mining is concerned with discovering qualitative and quantitative patterns in temporal databases or in discrete-valued time series (DTS) datasets. Recently two threads 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 and similar patterns in discrete-valued time series datasets (e.g., [3]), a general theory and general method of data analysis for discovering patterns from DTS is not well known. In this paper we describe a new framework for discovering patterns from temporal health records by using wavelet analysis and a data regression function. There are three steps. The first step of the framework employs of a distance measure and wavelet analysis to discover structural patterns (shapes). Coarse shapes of patterns are identified from the DTS and grouped into a wavelet model by the Nearest Neighbour (NN) algorithm employing a distance measure. In the second step the degree of similarity and periodicity between the extracted patterns is measured based on the data value distribution models. The third step of the framework consists of a hybrid model for discovering global patterns based on results of the first two steps.

The paper is organised as follows. Section 2 discusses related work. Section 3 describes our Wavelets Feature Model (WFM). Section 4 briefly explains the background of the application, describes the application of the approach to a real-world dataset and discuss the results. The final section concludes the paper with a brief summary.

2 Related Work

According to the principle of general pattern mining from a dataset we can classify objectives in pattern searching into three categories:

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

In recent years various studies have proposed algorithms for searching different kinds of and/or different levels of patterns. These studies have only covered one or sometimes two of the above categories. For example, most researchers use statistical techniques such as Metric-distance based techniques, Model-based techniques, or a combination of techniques (e.g, [8], [16]) to search different pattern problems such as in periodic pattern searching (e.g, [7, 9]) or in similar pattern searching (e.g, [5]).

Some studies have covered the above three categories for searching patterns in data mining. For instance, Agrawal et al. [1] presents a "shape definition language", called $\mathcal{SDL}$, for retrieving objects based on shapes contained in the histories associated with these objects. Das et al. [4] describes adaptive methods which are based on similar methods for finding rules and discovering local patterns and Baxter et al. [2] have considered three alternative feature vectors for representing variable-length patient health records.

In this paper we differentiate our approach in two ways. First, we use a statistical language to perform the search. Second, we divide the data sequence, or data vector sequence, into two groups: the structure based groups and the pure value based groups.

In the structure-based grouping our techniques are a combination of the work of Agrawal at al. [1] and Baxter et al. [2]. With this grouping we use a distance measuring function on the structural wavelet's based sequences, similar to the work of Berger in [14]. Alternatively we could use a model-based clustering method on the state-space $\mathcal{S}$ (such as Snob as used in [2]) to find clusters but it does not facilitate the understanding of the pattern distribution within the dataset.

In the value-based grouping we apply statistical techniques such as a frequency distribution function to deal with the actual values in relation to their structural distribution. This is similar to the work of Das et al. [4] but it benefits from combining significant information of the two groups to gather information underlying the dataset.

3 Wavelete's Feature-based Pattern Mining

This section presents our temporal data mining model in searching and analysing patterns from a DTS using the wavelet feature-based regression models (WFMs). For an analysis of a real-world temporal dataset which may contain different kinds of patterns, such as complete and partial similar patterns and periodic patterns, we consider two groupings of the data sequence separately. These two groupings are: (1) structure-based grouping and, (2) pure value-based grouping. For structural pattern search we consider the data sequence as a finite-state structural vector sequence applying a distance measure function in wavelet feature analysis. To discover pure value-based patterns we use data regression techniques on the data values. We then combine the results from both to obtain the final wavelet feature-based regression models (WFMs).

3.1 Definitions, Basic Models and Properties

We first give a definition of DTS and then provide some definitions and notation to be used later.

Definition 1 *Suppose $\{\Omega, \Gamma, \Sigma\}$ 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)****.*

We assume that for every successive pair of time points in the DTS t_{i+1} - t_i = $f(t)$ is a function (in most cases, $f(t)$ = constant). For every sequence of three time points: X_{i-1}, X_i and X_{i+1}, the triple (Y_{i-1}, Y_i, Y_{i+1}) has only nine distinct states (or nine local features), as enumerated in Figure 1, depending on whether the values increase, decrease or stay the same over the two time steps.

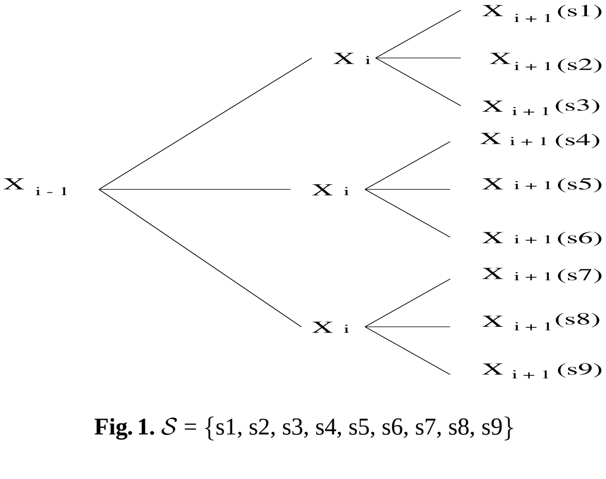

Fig. 1. $\mathcal{S}$ = {s1, s2, s3, s4, s5, s6, s7, s8, s9}

Definition 2 *In this framework suppose S_s represents the same state as the previous state, S_u represents an increase over the previous state, and S_d represents a decrease over the previous state. Let $\mathcal{S}$ = {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) }. Then $\mathcal{S}$ is the **state-space**.*

A sequence is called a *full periodic sequence* if every point contributes (precisely or approximately) to the cyclic behaviour 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 behaviour of the sequence is periodic at some but not all points in the time series.

Haar's Wavelet Function We choose the simplest basis function of the wavelet system—the *Haar wavelet basis function*—in this paper. Our use of Haar's wavelet function is limited to standard results taken from a well established literature. More details on the Haar wavelet function can be found in any standard wavelet textbook, including [15]. The Haar function was developed in 1910 [6] (and often called the $mother\ wavelet$) is given by:

$$\psi(x) = \begin{cases} 1 & \text{if } 0 \leq \text{x} < \frac{1}{2} \\ -1 & \text{if } \frac{1}{2} \leq \text{x} < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

If f is a function defined on the whole real line then for a suitably chosen $mother\ wavelet$ function ψ we can expand f as:

$$f(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} w_{jk} 2^{-j/2} \psi(2^{-j}t - k) \tag{2}$$

where the functions $\psi(2^{-j}t - k)$ are orthogonal to one another and w_{jk} is the discrete wavelet transform (DWT) defined as

$$w_{jk} = \int_{-\infty}^{\infty} f(t) 2^{-j/2} \psi(2^{-j}t - k) dt \tag{3}$$

where j and k are integers, j is a scale variable and k is a translation variable.

The Mahalanobis distance Function In a distributional space (e.g., state space, probability space), two conditional distributions with similar covariance matrices and very different means are so well separated that the Bayes probability of error is small. In this paper, we use Mahalanobis distance functions which are provided by a class of positive semidefinite quadratic forms. Specifically, if $\mathbf{u} = (u_1, u_2, \cdots, u_p)$ and $\mathbf{v} = (v_1, v_2, \cdots, v_p)$ denote two p-dimensional observations of each different distance of patterns in the same distributional space on objects that are to be assigned to two of the g pre-specified groups, then, for measuring the Mahalanobis distance between $\mathbf{u}$ and $\mathbf{v}$ we can consider the function:

$$D^2(i) = (\bar{\mathbf{u}} - \bar{\mathbf{v}})^{\mathbf{T}} \sum\nolimits^{-\mathbf{1}} (\bar{\mathbf{u}} - \bar{\mathbf{v}}) \tag{4}$$

where $\bar{\mathbf{u}}$ = $\mathbf{Eu}$, $\bar{\mathbf{v}}$ = $\mathbf{Ev}$ are means, and $\sum$ is a covariance matrix.

Local Linear Model We consider the bivariate data (X_1, Y_1), ...,(X_n, Y_n), which forms an independent and identically distributed sample from a population (X, Y). 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:

$$\mathbf{Y} = m(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon \tag{5}$$

where $E(\varepsilon)$ = 0, $Var(\varepsilon)$ = 1, and X and ε are independent. For an unknown regression function m($\mathbf{x}$), applying a Taylor expansion of order p in a neighbourhood of $\mathbf{x}_0$ with its remainder ϑ_p,

$$m(\mathbf{x}) = \sum_{j=0}^{p} \frac{m^{(j)}(\mathbf{x}_0)}{j!}(\mathbf{x} - \mathbf{x}_0)^j + \vartheta_p \equiv \sum_{j=0}^{p} \beta_j (\mathbf{x} - \mathbf{x}_0)^j + \vartheta_p. \tag{6}$$

The first stage of methods for detecting the characteristics of those records is to use linear regression. We may assume the linear model is $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$. The linear model based upon least square estimation (LSE) is $\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$. Then 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 ith diagonal element of $(\mathbf{X}^T\mathbf{X})^{-1}$.

3.2 Mining Global Patterns From a Database

For any dataset we divide the dataset into two parts: the qualitative part and the quantitative part. The qualitative part is based on the above state space for structural pattern searching and the quantitative part is based on probability space for statistical pattern searching.

We may view the structural base as a set of a vector sequences $\{\mathbf{S_1}, \cdots, \mathbf{S_m}\}$, where each $\mathbf{S_i} = (s_1, s_2, \cdots, s_p)^T$ denotes the p-dimensional observation on an object that is to be assigned to a prespecified group.

For qualitative pattern searching we first use multiresolution analysis (decomposition) with a Haar wavelet, then we apply the Mahalanobis distance functions on the state-space $\mathcal{S}_j = \{s_{1j}, s_{2j}, \cdots, s_{mj}\}$ of $\mathcal{S}$.

For quantitative pattern searching we only consider the structural relationship between the response variable Y and the vector of covariates $\mathbf{X} = (t, X_1, \ldots, X_n)^T$. By Taylor expansion we may fit a linear model as above and parameters can be estimated under LSE. The problem can then be formulated as the data distribution functional analysis of a discrete-valued time series.

We combine the above two kinds of pattern discovery to discover global information from a temporal dataset. For the structure group let the structural sequence $\{S_t : t \in \mathsf{N}\}$ be data functional distribution sequence on the state-space $\{s_1, s_2, \ldots, s_N\}$. Then suppose the pure valued data sequence is a nonnegative random vector process $\{V_t;\ t \in \mathsf{N}\}$ such that, conditional on $S^{(T)} = \{S_t :\ t = 1,\ \ldots,\ T\}$, the random vector variables $\{V_t :\ t = 1,\ \ldots,\ T\}$ are mutually independent.

4 An Application in Health Care Data

The dataset used in this study is Australian Medicare data. Medicare is the Australian Government's universal health care system covering all Australian citizens and residents. Each medical service performed by a medical practitioner is covered by the Medicare Benefits Scheme (MBS) and is recorded in the MBS database as a transaction. This dataset has been collected and stored since the inception of Medicare in 1975. Such a massive collection of data provides an extremely valuable resource of information. We present a case study on using our data mining techniques to analyse the medical service profiles of diabetes, a common disease in the senior population in Australia. This study complements a recent study of the time sequence dataset using a vector feature approach [2]. Similar to that study we use a subset of de-identified data (to protect privacy) based on Medicare transactions from Western Australia (WA) for the period 1994 to 1998. Our particular focus is on the patterns of care related to elderly diabetes patients (over 65 years of age).

Three monitoring medical tests for diabetes treatment are given in Table 1. These are essential for controlling the condition of diabetic patients.[1] Glycated hemoglobin measurements (*Gl*) provide information about the accumulated effect of glucose levels. Ophthalmologic examinations (*Op*) are important in the early identification and of complications related to eye sight. Cholesterol measurements via lipid studies (*Ch*) help identify possible complications relating to heart conditions.

Abbrev	Description	Guidelines
Gl	Quantitation of glycosylated hemoglobin.	2–4 times per year
Op	Ophthalmologic examination.	Every 1-2 years
Ch	Cholestorol measurement via lipid studies.	Every year

Table 1. Types of services received by Patients and indicative guidelines.

4.1 Experimental Results

The data used in this paper was extracted from the Medicare transactional database.[2] We used a subset of the de-identified data based on Medicare transactions from Western Australia (WA) for the period 1994 to 1998 inclusive. The sample data includes 4916 elderly diabetic patients. We have only limited demographic information about each patient, such as age, gender and location. For each patient we also have the sequence of diabetes-related monitoring tests they have received over the time interval. We identified clusters in which patterns associated with nine states for diabetes patients were found.

We study each of the three tests separately to find out how a patient's treatment follows the guidelines. We also study the overall patterns which take all three tests into consideration. To this end we summarise all the events (the tests performed) into eight distinct types of events which are listed in Table 2. A sample patient record is illustrated in Figure 2. For example, the patient has taken test 6 (i.e., Op and Ch) in early 1994, followed by a test 5 (i.e., Gl and Ch) and so on.

Through this experiment we are interested in investigating the following issues which are of interest to medical experts with particular interest in changes to patterns of care in the management of Diabetes over time:

[1] The Health Insurance Commission of Australia `http://www.hic.gov.au`.

[2] All experiments were done on a Unix system and under Windows NT with the prototype written in Awk and MATLAB.

Test group	Description	Test group	Description
test 0	No test	test 4	Gl and Op
test 1	Gl only	test 5	Gl and Ch
test 2	Op only	test 6	Op and Ch
test 3	Ch only	test 7	Gl, Op and Ch

Table 2. Eight possible test combinations tests for patients

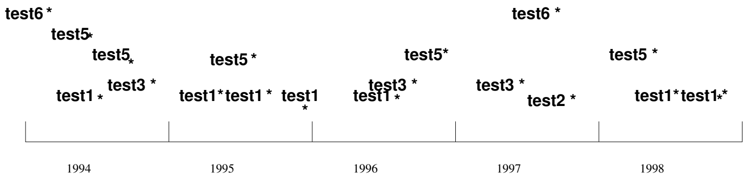

Fig. 2. A sample patient's health record, showing the seven types of tests received over five years.

- Does there exist any temporal pattern P_t for all patients who have one, two, or three tests regularly?
- What features are there for those temporal patterns? and
- Does there exist any temporal subpattern in P_t or between patterns P_t's?

Modelling DTS We assume that for each successive pair of time points in a DTS we have t_{i+1} - t_i = c (a unit constant). For mining temporal patterns from a real-world dataset we use *time gap* as the time variable between events (e.g., the same test group) instead of natural time (e.g., day) between different events. For example, a patient record is given in Table 3.

Test group	The day of test group taken	Time gap between the same test group
1	1	time gap = 1
1	191	time gap = 190
2	331	time gap = 1
7	487	time gap = 1
2	779	time gap = 448
1	894	time gap = 703
6	947	time gap = 1

Table 3. A patient test group transactional record with time gap.

From Table 3 we apply pattern searching on state-space and use time gap as variable for each test group for structural pattern searching. This means that we may view the structural base as a set of vector sequence: $\mathbf{S}_{9\times m} = \{\mathbf{S}_1, \cdots, \mathbf{S}_m\}$, where each $\mathbf{S}_i = (s1_i, s2_i, \cdots, s9_i)^T$ denotes the 9-dimensional observation on an object that is to be assigned to a prespecified group. Then the problem of structural pattern discovery for the sequence and its each subsequence $\mathbf{S}_{ij} = \{si_1, si_2, \cdots, si_j : 1 \leq i \leq 9, 1 \leq j \leq m\}$ of **S** on finite-state space can be formulated as a Haar's function with Mahalanobis distance model.

Then we may also view the value-point process data as N-dimensional data set[3]: $\mathbf{V} = \{\mathbf{V}_1, \cdots, \mathbf{V}_m\}$, where each $\mathbf{V}_i = (v1_i, v2_i, \cdots, vN_i)^T$, where the N is dependent on how many statistical values relate to the structural base pattern searching. Then the problem of value-point pattern discovery can be formulated as stochastic distribution of the sequence and its subsequences $\mathbf{V}_j = \{v1_j, v2_j, \cdots, vN_j\}$ of a discrete-valued time series [4].

On structural pattern searching We are investigating the data structural base to test naturalness of the similarity and periodicity on Structural Base distribution. We consider seven test groups in the state-space for structural distribution: $\mathcal{S} = \{s1, s2, \ldots, s9\}$. For finding all levels patterns(or clusters), we applied Haar's function (Equation 1) and distance function (Equation 4) on three-dimensional dataset for structural based pattern researching in state-space. First dimension is the test group, second dimension is the time gap between each test group and the third dimension is the time gap within the same test group.

For example for test group one and test group nine, in Figure 4, the *X*-axis represents the distribution frequency of test groups (e.g., in first dimension), the *Y*-axis represents distribution of time gap frequency between test groups (e.g., in second dimension) and the *Z*-axis represents distribution of frequency of time gap between the same test groups in state-space (e.g., in third dimension). Now we interpret an important result from the structural pattern searching: There exist similar time gap frequency patterns between and/or within state one and state nine for all test groups, this means, for example, pattern of patients not taking any test is similar to the pattern of patients taking test group one and test group two with the time gap increasing (or, decreasing).

We found some other results such as (1) there exist similar time gap frequency patterns between and/or within state two and state six for all test groups: the meaning is that the patients have not been given good care by doctors according to the clinical guidelines and (2) there exist similar and periodic time gap frequency patterns between state three and state seven for all test groups: the reason for the pattern is that the patients have been given good care by doctors according to the clinical guidelines.

In Figure 5, the *x*-axis represents natural integer sequence $\mathbf{N}$ and the *y*-axis represents the time gap for each state. Figure 5 explains some important facts: first that there exists the same time gap statistical distribution (e.g., the same tangent curve distribution) between the test group 1, test group 2, test group 3 and test group 5. It also explains visits to doctors are stationary in different time gaps for those four types of test groups. Second, there exists a hidden periodic distribution which corresponds to patterns on the same state with different distances, this means patients visit their doc-

[3] According to their structural distribution model

[4] In fact, many practical problems in temporal data mining related to statistical modelling are explained in the context of regression models.

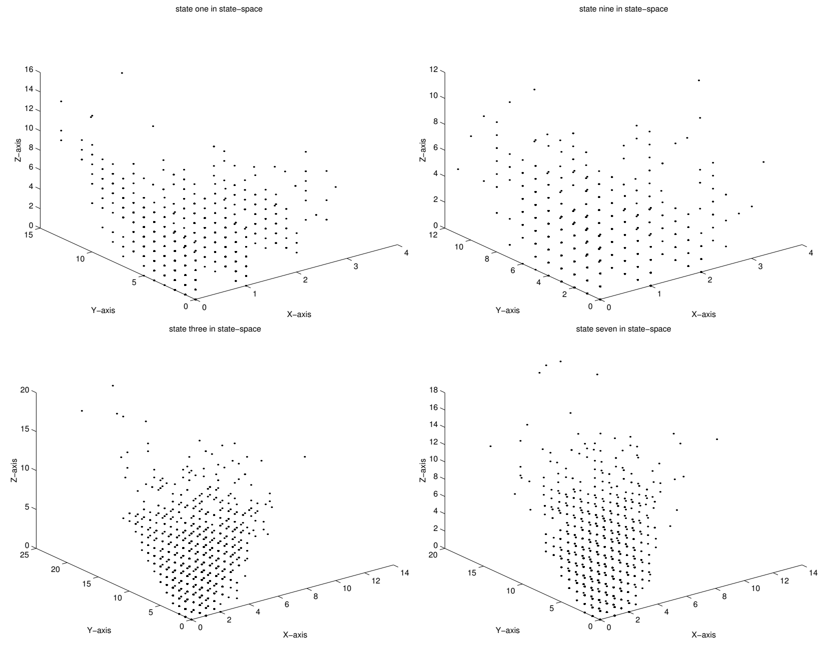

Fig. 3. All test groups of all patients in state-space one s1 and nine s9, state-space three s3 and seven s7.

tors periodically and third there exist partial periodic patterns on and between some test group in state-space.

On value-point pattern searching We now illustrate our new method to analyse the value-point sequence of health temporal records for searching patterns. In these records, since each patient record length is different, we can only use their statistical value as variables in regression functions (e.g., frequency distribution functions). In the light of our structural base experiments, we have the series

$$Y_t = f_t^{testgroupi}(v_t) - f_t^{testgroupj}(v_t) \tag{7}$$

where $f_t^{testgroupi}(v_t)$ is a frequency distribution function, its variable v_t is the time gap between the same state (e.g., v_t = $state\ k_{t_1}$ - $state\ k_{t_2}$), in the same cluster. Then the observations can be modelled as a linear regression function,

$$Y_t = f_t^{testi}(v_t) - f_t^{testj}(v_t) + \varepsilon_t, \qquad t = 1, 2, \ldots, N \tag{8}$$

and we also consider the $\varepsilon(t)$ as an auto-regression $AR(2)$ model

$$\varepsilon_{t'} = a\varepsilon_{t'-1} + b\varepsilon_{t'-2} + e_{t'} \tag{9}$$

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.

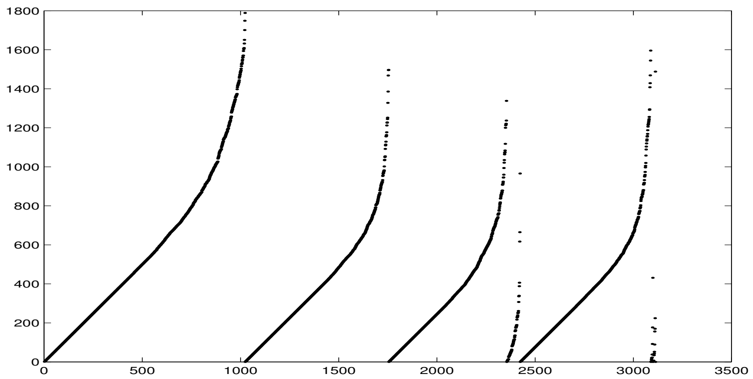

Fig. 4. Plot of the time gap within each test group for all 8 test groups in 1,875 business days.

In top of Figure 6 the *x*-axis represents the frequency of time gap between the same states and the *y*-axis represents the time gap between the same states k. This explains two facts: (1) there exists a Poisson distribution for each of test group 1 and test group 5. This means that period of the medical treatment for test group 1 and test group 5 is stationary independent increment. And (2) there exist the same patterns between two test groups with small distance shiftting, this means that patients have received treatment for test group 1 or test group 5 by the guidelines.

Left hand of bottom of Figure 6 shows that there exists an exponential distribution for test group 2. This means that the patient has a problem and is receiving treatment for the identification or control of the problem and in botton of Right hand of bottom of Figure 6 shows that there exists a geometric distribution for test group 3. This means that patients have received a regular treatment.

4.2 Mining Global Patterns

According to the above analysis in the health data record, let $\{S_t : S_t \in \mathcal{S}, t \in \mathsf{N}\}$ be a structural process representing $state\ k$ occurrence, and $\{V_t : t \in \mathsf{N}\}$ be the corresponding observed values, then we have the distribution of V_t conditional on S_t given by

$$\mathsf{P}(V_t = v | S_t = i) = p_{vi}^t \tag{10}$$

For test group 1 and test group 5, $V_t^{testgroup1}$ and $V_t^{testgroup5}$ both have Poisson distribution with means $\lambda_i^{testgroup1}$ and $\lambda_i^{testgroup5}$. Two states satisfy

$$V_t^{testgroup1} = \alpha V_t^{testgroup5} + \theta_t \tag{11}$$

Then the conditional mean of V_t and state-dependent probabilities given for all non-negative integers v_t will be (it is the same as 7, 8)

$$\begin{aligned} &\mu(t) = \textstyle\sum_{i=1}^{m} \lambda_i W_t(t), \\ &P_{v_t,statek} = e^{-\lambda_{i,v_t}} \frac{\lambda_{i,v_t}}{v_t!} \end{aligned} \tag{12}$$

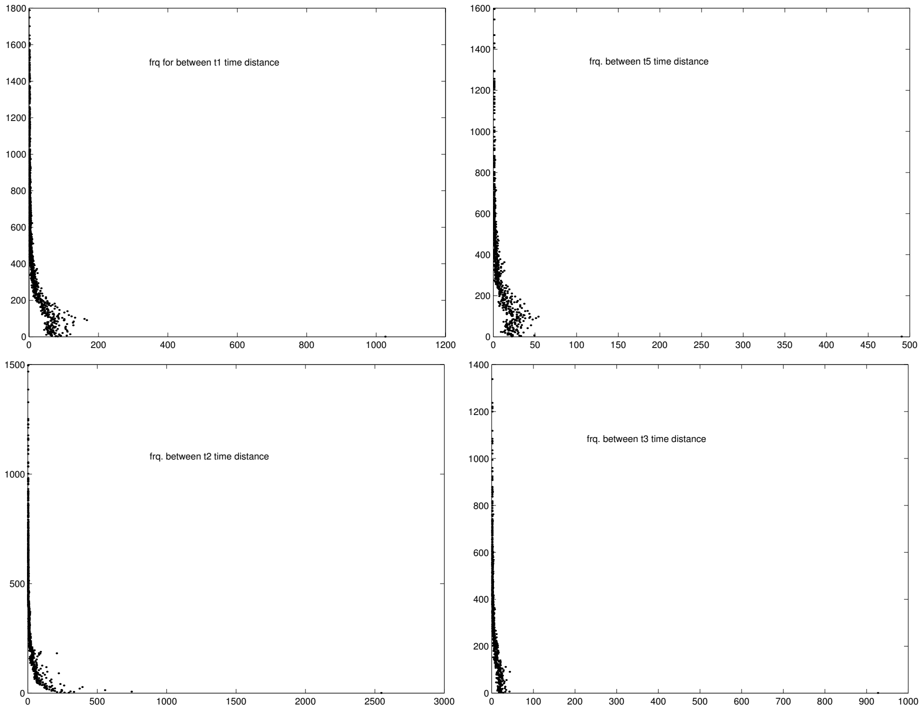

Fig. 5. Top: Time distance frequency distribution of test group 1 and test group 5 are Poisson distributions. Bottom: Time distance frequency distribution of test group 2 is an exponential distribution (Left) and Time distance frequency distribution of test group 3 is a geometric distribution (Right).

For test group 2, $V_t^{testgroup2}$ is an exponential distribution with parameters $\lambda_i^{testgroup2}$ and $\mu_i^{testgroup2}$. Then the conditional exponential distribution of $V_t^{testgroup2}$ and state-dependent probabilities given for all non-negative integers v_t will be

$$\begin{aligned} &\mu(t) = \textstyle\sum_{i=1}^{m} \lambda_i W_t(t), \\ &P_{v_t,test2} = \begin{cases} \lambda_{(i,v_t)} e^{(-\lambda_{(i,v_t)}(v_t-\mu))} & v_t > \mu \\ 0 & v_t < \mu \end{cases} \end{aligned} \tag{13}$$

For test group 3, $V_t^{testgroup3}$ is a geometric distribution with parameter $p_i^{testgroup3}$. Then the conditional geometric distribution of $V_t^{testgroup3}$ and state-dependent probabilities given for all non-negative integers v_t will be

$$\begin{aligned} &\mu(t) = \textstyle\sum_{i=1}^{m} \lambda_i W_t(t), \\ &P_{v_t,testgroup3} = p_{i,v_t}^{statek}(1 - p_{i,v_t}^{statek})^{(v_t-1)} \end{aligned} \tag{14}$$

Test group 4, test group 6 and test group 7 are indepentent states. We use Haar function and local polynomial function for each of them to find their conditional distribution function $f_{testgroupk}(t)$. We found that there exist some similar patterns between each of their state but no patterns exist between each of their clusters of time gap, this means the patients have received number of treatments from test group k (k = 4, 6, 7) similar but for different time periods.

The main results from structural pattern seaching and value-point pattern searching are:

- The behaviour of visiting doctors for pattients with just a diabetic is a Poisson distribution, the meaning is that number of doctor visits by diabetic group people in the same of period(e.g., within 7 days) have the same distribution.
- The distribution of visiting pattern(taking tests) between the patient that has taken more care and less care is log-normal distribution, the meaning is that the number of doctor visits is not symmetric around the mean, but much extending to the right(e.g., pattern of less care)
- The other main combined-results on the health dataset are as follows:
  1. There does exist some full periodic pattern within and between state 1 and state 5, this means the time gap between patients taking test 1, and time gap between patients taking test 1 & test 3 are both stationary.
  2. There exist some partial periodic patterns between state 1, state 2, state 3 and state 5, this means the patients have sub-common problem such as they all have eye problem(e.g., taking more eye test, etc.).
  3. There also exist some similar patterns between state 1, state 2, state 3 and state 5. This means there exists similar patterns of behaviour for patients visiting their doctors but for different tests.

In [2] three alternative feature vectors for representing variable-length patient health records were used. An interesting observation from there is that the average, residual, and deviance clusters have similar mean average and deviance values, but differ in their residual value. This shows the value of using the residual feature to identify intensive patterns of care during a relatively short time interval. In our approach, using a temporal data mining method, we can find results (e.g., interesting patterns and unexpected patterns).

5 Concluding Remarks

This paper has presented a new approach based on hybrid models 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 similar patterns. The clusters of similar 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 wavelet's feature-based regression models (WFMs) 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 in [10, 11, 13, 12]; there the models used are based on hidden functional analysis. However, we have found that using different models at different levels produces better results.

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.

References

1. R Agrawal, G Psaila, E L Wimmers, and M Zait. Querying shapes of histories. In *Proceedings of the 21st VLDB Conference*, 1995.
2. R A Baxter, G J Williams, and H He. Feature selection for temporal health records. In *Proceedings of the 5th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2001.
3. C Bettini. Mining temportal relationships with multiple granularities in time sequences. *IEEE Transactions on Data & Knowledge Engineering*, 1998.
4. G Das, K Lin, H Mannila, G Renganathan, and P Smyth. Rule discovery from time series. In *Proceedings of the international conference on KDD and Data Mining(KDD-98)*, 1998.
5. G Das, D Gunopulos, and H Mannila. Finding similar time seies. In *Principles of Knowledge Discovery and Data Mining '97*, 1997.
6. A Haar. Zur theore der orthoganalen funktionen systeme. *Annals of Mathematics 69: 331-371*.
7. J Elder IV and D Pregibon. A statistical perspective on knowledge discovery in databases. In U Fayyad, G Piatetsky-Shapiro, P Smyth, and R Uthurusamy, editors, *Advances in Knowledge Discovery and Data Mining*, pages 83–115. The MIT Press, 1995.
8. J B MacQueen. Some methods for classification and analysis of multivariate observations. In *5th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, 1967.
9. C Li and G Biswas. Temporal pattern generation using hidden markov model based unsuperised classifcation. In *Proc. of IDA-99*, pages 245–256, 1999.
10. W Q Lin and M A Orgun. Applied hidden periodicity analysis for mining discrete-valued time series. In *Proceedings of ISLIP-99*, pages 56–68, Demokritos Institute, Athens, Greece, 1999.
11. W Q Lin and M A Orgun. Temporal data mining using hidden periodicity analysis. In *Proceedings of ISMIS2000*, University of North Carolina, USA, 2000.
12. W Q Lin, M A Orgun, and G J Williams. Temporal data mining using hidden markov-local polynomial models. In *Proceedings of PAKDD2001*, The University of Hongkong, Hong Kong, 2000.
13. W Q Lin, M A Orgun, and G J Williams. Temporal data mining using multilevel-local polynomial models. In *Proceedings of IDEAL2000*, The Chinese University of Hongkong, Hong Kong, 2000.
14. S Jajodia, S Sripada, and O Etzion, editors. *Temporal databases: Research and Practice*. Springer-Verlag, *Lecture Notes in Computer Science*, Volume 1399, 1998.
15. H L Resnikoff and R O Wells, editors. *Wavelet Analysis, The scalable structure of information*. Springer-Verlag, 1998.
16. Z Huang. Clustering large data set with mixed numeric and categorical values. In *Proceedings of the 1st Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 1997.