Multilevels Hidden Markov Models For Temporal Data Mining

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 paper describes new temporal data mining techniques for extracting information from temporal health records consisting of time series of diabetic patients' treatments. In this new method, there are three steps for analyzing patterns from a longitudinal data set. The first step, a structural-based pattern search, to find qualitative patterns (or, structural patterns). The second step performs a value-based search to find quantitative patterns. In the third step we combine results from the first two steps to form new model. The hidden Markov model has the expressive power of both qualitative analysis and data quantitative analysis. The global patterns can therefore be identified from a DTS set.

Keywords: temporal data mining, discrete-valued time series, similarity patterns, periodicity analysis, hidden Markov model

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). Recently, there are 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 in discrete-valued time series (DTS) datasets (e.g. [1]), a general theory and general method of data analysis of discovering patterns for DTS data analysis is not well known. In this paper we describe a new framework for discovering patterns from temporal health records using multilevel hidden Markov model(MHMM). There are three steps for discovering knowledge from the dataset in this approach. The first step of the framework consists of a Markov model analysis for discovering structural (qualitative) patterns. In this step, only the rough shapes of patterns are decided from the DTS. The patterns are grouped into clusters by Nearest Neighbour (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 the data statistical distribution (quantitative patterns). The third step of the framework consists of a Hidden Markov model for discovering global patterns based on results of the first two steps.

The paper is organised as follows. Section 2 presents the medical data background, definitions, basic methods and our new method of pattern discovery. Section 3 applies new models to real-world dataset "diabetes dataset". The last section concludes the paper with a short summary.

2 The Problem Background, Definitions and Basic Methods

2.1 Health Data Background

Medicare is the Australian Government's universal health care system. Each visit to a medical practitioner or hospital is covered by Medicare and recorded as a transaction in the Medicare Benefits Scheme (MBS) database. This data has been collected in Australia since the inception of Medicare in 1975. Such a massive collection of data provides an extremely rich resource that has not been fully utilised in the exploration of health care delivery in Australia. The HIC has a responsibility to protect the public purse and to ensure that taxpayer's funds are spent wisely and efficiently on health care. The knowledge discovered can be used to educate medical practitioners to improve their medical practice in order to achieve the best health outcomes while ensuring health costs remain under control. In this paper, we present our new temporal data mining techniques to analyze the medical service profiles of diabetes, a relatively common disease amongst the senior population in Australia.

For this current exploration we use a subset of de-identified data (to protect privacy) based on Medicare transactions for the period 1997 to 1998. Our particular focus is on patterns of care related to diabetes. For example, we can ask ourselves questions like: Are there any distinct patterns of care for these diabetes patients? Are there any groups of patients receiving similar patterns of care? Are the patterns of care related to their doctor? Do patients of different ages or gender or location receive differing patterns of care to other patients? Answers to the above questions rely on a thorough analysis of the sequences of medical test of the patients and is the objective of our research.

We first give a definition for what we mean by DTS and background of health care data. We then give some definitions and notations which will be used later. We will also explain the basic models briefly. The detail of the models will be illustrated 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* [1] *(DTS).*

[1] Many 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 instance, most DTS data sets can be represented by means of models based on normal distribution.

2.2 Definitions and Properties

We consider the bivariate data (X_1, Y_1), …,(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), $i = 1, 2, \ldots, N$, we can regard the data as being generated from the model

$$\mathbf{Y} = m(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon \tag{1}$$

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 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, e.g., Figure 1).

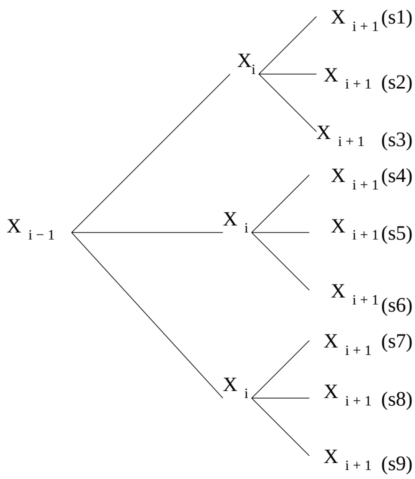

Fig. 1. For every successive three points in nine states space: **S** = {s1, s2, …, s9}

Definition 2 *Suppose S_s is the same state as prior one, S_u is the go-up (or, stonger) state compare with prior one and S_d is the go-down (or, weaker) state compare with prior one. Let **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 **S** called state-space. If S_s, S_u and S_d are all indepentent state to each other, then **S** = {1, 2, …, N} also called state space.*

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.

Definition 3 *Let $h = \{h_1, h_2, \ldots\}$ be a sequence. If for every $h_j \in h$, $h_j \in$ **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 4 *Let $y = \{y_1, y_2, \ldots.\}$ be a real value sequence, then y called a value-point process. If 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. In general, if the sequence y_j has $h(t) - \delta < y_j < h(t) + \delta$ for all j, we say that y has an approximate distribution function h(t).*

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.

2.3 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 also limited to standard results taken from the literature. We have in particular used those of Baldi and Brunak [9].

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:

$$\eta_{ij} = \mathsf{P}(\ S_t = j \mid S_{t-1} = i\ ) \tag{2}$$

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

$$\mathsf{P}(\xi_t = v | S_t = i) = \pi^t_{vi} \tag{3}$$

where the probabilities π^t_{vi} as the "state-dependent probabilities [2]". If the probabilities π^t_{vi} do not depend on t, the subscript t will be omitted.

[2] The models $\{\xi_t\}$ are defined as hidden Markov 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" $\{S_t\}$.

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. Its autocorrelation function ρ_k of S_t:

$$\rho_k = \frac{Cov(S_t, S_{t+k})}{VarS_t} \tag{4}$$

A useful device for depicting the dependence structure of such a model is the conditional independence graph. Figure 2, displays the independence of the observations $\{C_t\}$ given the states $\{V_t\}$ occupied by the Markov chain, as well as conditional independence of C_{t-1} and C_{t+1} given C_t.

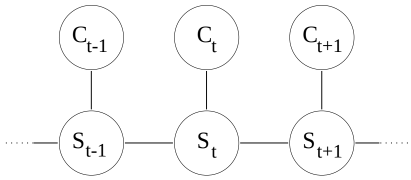

Fig. 2. Conditional independence graph of hidden Markov model.

2.4 Pattern Discovery

In this subsection, we apply our new data mining model for discovering qualitative and quantitative temporal patterns analysis in a DTS by hidden Markov model[3].

For building up our new data mining model, we consider two groups of the data sequence separately. These two groups are: (1) structural based data group and, (2) pure value-based data group. In group one, we only consider the data sequence as a finite-state structural vector sequence, applying Markov model for performing qualitative pattern search. In group two, we use data regression function on pure value-based sequence data for discovering quantitative temporal patterns. Then we combine those two groups by using HMM analysis to obtain final results.

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). According to our new method, We may view the structural base as a set of vector sequence: $\mathbf{S}_{9\times m} = \{\mathbf{S}_1, \cdots, \mathbf{S}_m\}$ in function(1), 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 Markov model.

[3] For more theory and applications, see journals such as *IEEE Transactions on Signal Processing* and *IEEE Transactions on Speech and Audio Processing.*

Then we may also view the value-point process data as N-dimensional data set, according to their structural distribution: $\mathbf{V} = \{\mathbf{V}_1, \cdots, \mathbf{V}_m\}$, 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 distributon of the sequence and its subsequences $\mathbf{V}_j = \{v1_j, v2_j, \cdots, vN_j\}$ of a discrete-valued time series [4].

Structural Pattern Discovery We now introduce an approach to discovering patterns in structural base vector sequences by a Markov model. We apply Markov model on state-space by using equation 2 and $\rho_k = \rho_1^k$ for all $k \in \mathbf{N}$ for each subsequence $\mathbf{S}_j = \{s_{1j}, \cdots, s_{Nj}\}$. Second step is that finding each of sub-transition probability matrix $\varDelta_j$ vales by using the limit theorem and final the structural distribution for each of them (clusters) can be decided.

Point-Value Pattern Discovery On the value-point pattern discovery, we only consider the structural relationship between the response variable Y and the vector of covariates $\mathbf{X} = (t, X_1, \ldots, X_n)^T$. For a given dataset, the unknown regression function (in equation 2 [5]) m($\mathbf{x}$), applying a Taylor expansion of order p in a neighbourhood of $\mathbf{x}_0$ with its remainder ϑ_p,

$$m(\mathbf{x}) = \sum_{j=0}^{p} \frac{m^{(j)}(\mathbf{x}_0)}{j!}(\mathbf{x} - \mathbf{x}_0)^j + \vartheta_p \equiv \sum_{j=0}^{p} \beta_j (\mathbf{x} - \mathbf{x}_0)^j + \vartheta_p. \qquad (5)$$

The first stage of method for detecting the characteristics of those records is to use the linear regression analysis. We may assume 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 $(\mathbf{X}^T\mathbf{X})^{-1}$.

Now, for each value-point, we may fit a linear model as above and parameters can be estimated under LSE. Then the problem can be formulated as the data distribution functional analysis of discrete-valued time series.

Mining Global Patterns From a Dataset We combine the above two kinds of pattern discovery in a single framework. We introduce an enhancement to the hidden Markov modelling approach through functional data analysis.

In structure group, let the structural sequence $\{S_t : t \in \mathsf{N}\}$ be an irreducible homogeneous Markov chain on the state space $\{s1, s2, \ldots, s6\}$, with the transition probability matrix $\varDelta$ [6].

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

[5] We always denote the conditional variance of Y given $\mathbf{X} = \mathbf{x}_0$ by $\sigma^2(\mathbf{x}_0)$ and the density of $\mathbf{X}$ by $f(\bullet)$

[6] In fact, the transition probability matrix $\varDelta$ are all different for each classes and nonnegative random process $\{V_t;\ t \in \mathsf{N}\}$ as well for each classes.

In pure value group, suppose the pure valued data sequence is a nonnegative random vector process $\{V_t;\ t \in \mathsf{N}\}$ such that, conditional on $S^{(T)} = \{S_t : t = 1, \dots, T\}$, the random vector variables $\{V_t : t = 1, \dots, T\}$ are mutually independent and, if $S_t = I_{i\times i}$, V_t takes the value v with probability π^t_{vi} for each subsequence. For example, let $V_t = i$, ξ_t has a statistical distribution with parameters $n_{p,t}$ (a known positive integer) and p_i. That is, the conditional statistics distribution of ξ_t has parameters $n_{p,t}$ and $m(t)$, where

$$m(t) = \sum_{i=1}^{m} p_i W_i(t), \tag{6}$$

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

For examples, if $S_t = i$, ξ_t has a binomial distribution with parameters n_t (a known positive integer) and p_i. That is, the conditional binomial distribution of ξ_t has parameters n_t and $p(t)$, where

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

and $W_i(t)$ is, as before, the indicator of the event $\{S_t = i\}$. Then for $v = 0, 1, \dots, n_t$:

$$\pi^t_{vi} = \binom{n_t}{v} p_i^v (1 - p_i)^{n_t - v}$$

The models $\{\xi_t\}$ are defined as Binomial hidden Markov 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 Experimental Results

The data used in this case study was extracted from the Medicare transactional database[7]. We used the patients' Medicare services data from 1997 to 1998 for two calendar years. The sample data includes 10,000 diabetic patients. The data extracted from Medicare is raw transaction data which is stored on IBM main 370 frame computer running MVS operating system. It is a very large data set with millions of records and each record has more than a hundred attributes. There is a transaction record for each Medicare service. Each service record has its item number which is the most important field in the data set. The item number tells to a large extent what kind of service has been performed on the patient. The patients' medical service pattern is represented by a series of item numbers served during the year. The sample record structure is as follows:

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

@+q	....	Encrypted Provider number
766666769	....	Encrypted PIN number
6	....	Method of Payment
66341	....	Item number
14OCT1997	....	Date of service
16.85	....	Benefit
b	....	Reason code for rejection
79+@	....	Referal provider
0	....	Processing indicator
14OCT1997	....	Date of referal
a	....	Hospital index

In this experiment, we use our new method to analyze the medical service profiles of diabetes. We have applied our technique to identify clusters, in which the pattern of the diabetic patients were found. A sample of four patient records is illustrated in figure 3. The event sequence data can be augmented with any available vector based data.

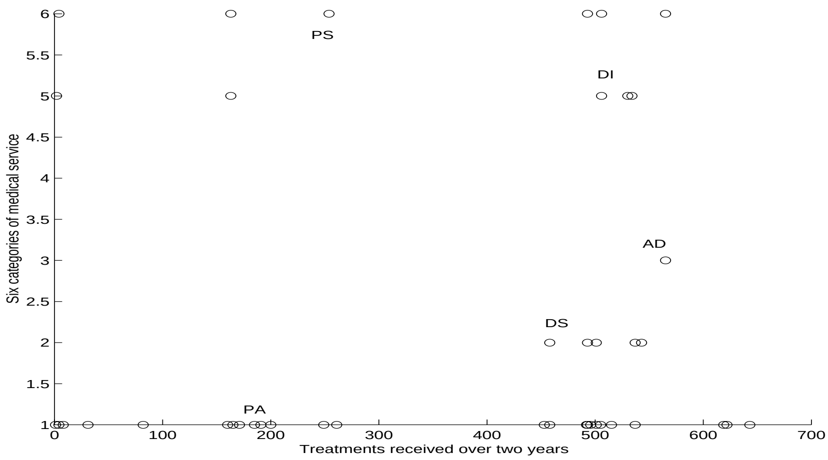

Fig. 3. A sample of a patient's health record, showing that the six categories of medical services have been received over two years.The medical services are: professional Attendances (PA); Diagnostic Services (DS); Approved Dental Practitioner Services (AD); Diagnostic Imaging Services (DI); and Pathology Services (PS).

There are three steps of experiments for the investigation of "Patterns in Diabetese-dataset" between the diabetic patients. Through this experiment, we are interested in finding following clusters and patterns:

- Does there exist any temporal pattern P_t for all types diabetic patients who do have visits to doctors (or, visiting pattern).
- Does there exist any temporal pattern within P_t (or, diffenet visiting subpattern for each kind of diabetic patients)?
- How many classes(clusters) are there in the dataset?

– What kinds of patterns (models) are there for each class?
– What kind of a relationship exists between classes?

3.1 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 6 states in the state-space of structural distribution: **S** $= \{s1, s2, s3, s4, s5, s6\}$.

For searching structural (or, qualitative) patterns, we use limit theorem on transition probability matrix Δ_k for each patient. We found there are exist three clusters:

– In cluster one, visiting temporal pattern is a stationary process. The meaning is that all the patients see their doctors regularly for consulatations and treatments,
– In cluster two, visiting temporal pattern is combinatoral Poisson process. This mean is that all the patients temporal behaviour(e.g., time distance between successive pair of two consultations or, between successive pair of two treatments) not symmetricaround their mean, but much extending to the right(e.g., patient has a fewer number of consulatations than the number of medical-treatments). It is also tell us, the patient in this group is a new diab-patient, or the patient has other problems other than diabetes, so they need more medical treatments.
– In cluster three, all the patients have a larger number of consulatations than the number of medical-treatments, for example, they may get other sicknesses such as cold very easily.

For example, in Figure 4, figures in each row from bottom to top is cluster one, cluster two and cluster three. “CC” means consultation to consultation, “CT” means consultation to treatments and so on. The *x*-axis represents the transition probability before using the limit theorem and the *y*-axis represents transition probability after using the limit theorem. This explains two facts: (1) there exists some sub-group which corresponds to patterns on the same condition in the group, and (2) there exist partial similar patterns between clusters. In summary, some results for the structural base

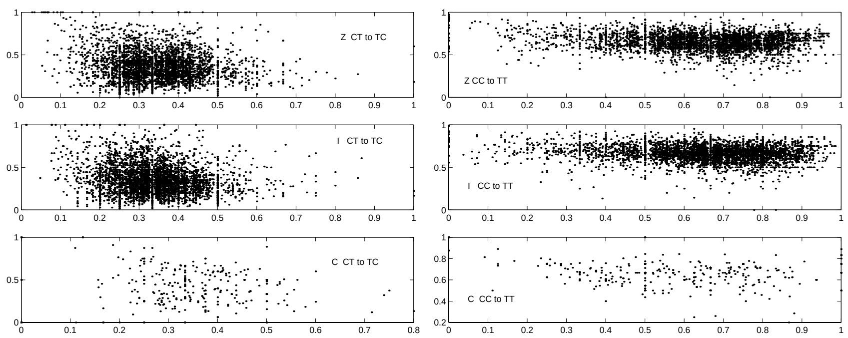

Fig. 4. plot of the transition probability between consulatation state and medical-treatment state in all states.

experiments are as follows:

- more than half percent of diabetes-patients have other major medical problems: patients have other major medical problems based on their type of diabetes. There exist similar patterns between major medical problems in structure of each of three clusters.
- less than half percent of diabetes-patients get “seasonal health problems” very easily: there also exist some similar patterns between each subgroup in structure (it may be related to their age).
- only 0.05 percent of normal patients are diab-patients: this means they do not have other major medical problems and they do not easily get “seasonal health problems”.

3.2 On value-point pattern searching

We now illustrate our new method on the value-point sequence for searching patterns. Recall data model and data vector sequence

$$\mathsf{Y} = m(\mathbf{V}) + \sigma(\mathbf{V})\varepsilon = \sum_{j=0}^{p} \beta_j(\mathbf{v} - \mathbf{v}_0)^j + \vartheta_p, \qquad let\ \mathbf{V}_j = \{v_{1j}, v_{2j}, \cdots, v_{Nj}\},$$

each vector $\mathbf{V}_j$ are time vector variable, e.g., v_{kj} (k =1, 2, . . ., N) is the time length between the state sk and sj has occurred.

In the light of our structural base experiments, we have found that there exists a linear regression function between categories [8] in each of the three clusters. For example, if let $C_i^k(t)$ be a time series distribution function for one category k, and its response variable is time distance between successive two time point(e.g., a patient visits doctor and use items in same category) then the time series $V_{i,j}^k(t)$ which difference between two linear regrassion functions at time t values $V_{i,j}^k(t) = C_i^k(t) - C_j^k(t)$ is an approximate distribution function.

For example, in Figure 5, the *x*-axis represents how many items have been used, the *y*-axis represents the accumulated occurences of each item, and the *z*-axis represents six categories.

Some results for the value-point of experiments are as follows [9]:

- There does exist some periodic patterns in the same cluster but with different category.
- There exist some similarity patterns in each cluster, but between different categories.
- There exist some similarity patterns between the same category from each cluster.

[8] There are six categories in Australia Medicare Benefits Schedule Book
[9] We classify repeating patterns based on a distance classification technique.

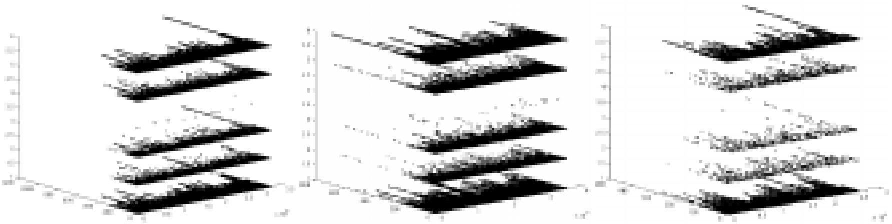

Fig. 5. Plot of the accumulated frequency count of each of the six categories in each of three clusters for 10,000 patients.

3.3 Using HMM for Global Pattern Searching

Let structural sequence $\{S_t : t \in \mathsf{N}\}$ be an irreducible homogeneous Markov chain on the state space $\{s1, s2, \ldots, s6\}$, with the transition probability matrix Δ.

In this experiment, the S_t and Δ are structural vector sequence. There are three different S_t and Δ according to their have different three groups (clusters), say, S_t^1, Δ^1, S_t^2, Δ^2, and S_t^3, Δ^3.

We are interested in the future of distribution of TPM, $f(t) = \Delta^t$ for each group. According to the Markov property, the TPM:

$$\lim_{t \to \infty} \Delta^t = \begin{cases} constant & \text{cluster one} \\ 0 & \text{cluster two} \\ \infty & \text{cluster three} \end{cases}$$

We have three value-point sequence V_t^1, V_t^2 and V_t^3 are nonnegative random processes from each cluster such that, conditional on S_t^1, S_t^2 and S_t^3 for each cluster and each of them also satisfy $Z_t = \alpha Z_{t-1} + \theta_t$.

Then for each group, the distribution of sequence of transition probability matrix(TPM) Δ^i under time order $\Delta_1^k, \Delta_2^k, \cdots, \Delta_t^k; t \in \mathsf{N}, k \in \{1, 2, 3\}$ corresponding to the prediction value-point $V_{i,j}^k$.

We have main combined-results on diabetic dataset as follows:

- There does exist three identifiable different groups(clusters). The data of the diabetic dataset only can be assigned to one of three prespecified groups based on their qualitative structure.
- There does exist some full periodic patterns within each category and some partial periodic patterns as well.
- There exist some similarity patterns between clusters, e.g., they have similarity patterns for visiting doctors.
- There does exist some subpatterns within each category in a cluster. This indicates there are some different types of diabetes, e.g., eye problem with diabetes, heart problem with diabetes.

4 Related Work

In recent years various studies have considered temporal datasets for searching different kinds of and/or different levels of patterns. For example, many researchers use statistical techniques such as Metric-distance based techniques, Model-based techniques, or a combination of techniques (e.g, [10]) to search for different pattern problems such as in periodic patterns searching (e.g., [4]) and similarity pattern searching (e.g., [3]).

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 [8]. In group two we apply statistical techniques such as local polynomial modelling to deal with pure data which is similar to [2]. Finally, our work combines significant information of two groups to get global information which is behind the dataset.

5 Concluding Remarks

This paper has presented a new approach of hidden Markov models to mining temporal patterns. 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 on state-space by the choice of certain states. 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 HMM model to obtain a global pattern picture and understand the patterns in a dataset better. Another approach to find similar and periodic patterns have been reported in [5–7]; there the model used is based on hidden periodicity 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 have been shown to be promising and we are currently applying the method to other types of large realistic medical data sets.

References

1. C. Bettini. Mining temportal relationships with multiple granularities in time sequences. *IEEE Transactions on Data & Knowledge Engineering*, 1998.
2. 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.

3. G.Das, D.Gunopulos, and H. Mannila. Finding similar time seies. In *Principles of Knowledge Discovery and Data Mining '97*, 1997.
4. Cen Li and Gautam Biswas. Temporal pattern generation using hidden markov model based unsuperised classifcation. In *Proc. of IDA-99*, pages 245–256, 1999.
5. Wei Q. Lin and Mehmet 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.
6. Wei Q. Lin and Mehmet A.Orgun. Temporal data mining using hidden periodicity analysis. In *Proceedings of ISMIS2000*, University of North Carolina, USA, 2000.
7. 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.
8. S.Jajodia andS.Sripada O.Etzion, editor. *Temporal databases: Research and Practice*. Springer-Verlag,LNCS1399, 1998.
9. P.Baldi and S. Brunak. *Bioinformatics & The Machine Learning Approach*. The MIT Press, 1999.
10. Z.Huang. Clustering large data set with mixed numeric and categorical values. In *1st Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 1997.