An Inner Distance Combination Transform for Classification of Temporal Medical Data

Klaokanlaya Silachan, Panjai Tantasanawong

Abstract

The classification of temporal medical data has received considerable attention, because it can reveal useful information hidden within a database. The objective of models for temporal medical data is to increase the classification accuracy. However, the problem with classifying temporal medical data for each patient is that we cannot deal with multiple values for each measurement, and classifiers cannot deal with multiple classes for each patient. This is because measurements in patient case series consist of a sequence of measurements taken at ordered points in time. This paper presents a novel inner distance combination transform method that converts temporal medical data into a new feature values that is then fed into a naïve Bayes classifier. We compare the new transform method with mean, median, mode and the final treatment result. The experimental results show that the proposed method is more accurate than the other methods.

Keywords: Feature transform, Temporal medical data, Temporal classification

1. Introduction

The mining of temporal medical data has been receiving considerable attention, as it provides a way of revealing useful information hidden within a database. Mining and the analysis of temporal data can reveal important patterns, and are important in eliciting further knowledge and utility from the data. In recent years, researchers have successfully identified temporal patterns by mining temporal data [4] and the time-related features of the data [14, 20]. Temporal classification is challenging, and successful classification is reliant on accuracy. The temporal data used for the analysis must be used to build the classification model. A learning model uses an actual set of historical data. These data describe measurements taken over a period of time, and the resulting predictions are the measurements expected to occur in the future [11, 19, 16]. However, one issue of particular interest in this area is the analysis of temporal data. The objective of classification models is to provide accurate predicted measurements. Temporal data contain valuable information about patient health and are extracted over a period of time [16, 17, 21]. Temporal medical data can be defined as a sequence of measurements taken at ordered points in time. These data are patient case series, and may refer to short-term and long-term treatments [13].

The goal of temporal data classification is to predict which class a given sequence belongs to, based on a set of training instances represented the feature values. The learning of a function assigns data records with certain attribute values into a predefined class [6, 17, 18]. The task of pattern classification is to map a given set of series to one of the predefined classes. Each event occurring at each time point has a value composed of continuous observations of interest, and each measurement consists of multiple time values [1] [10].

We must consider the problem of building a classifier to predict patient measurements using past patient data. Unfortunately, to enable the accurate classification of a new patient with similar data to that in the training set, we must know the classification accuracy of the new feature values. When training the classifier directly, it is useful to have collections of temporal medical data from the patient case series for the model to learn. However, it is very difficult to examine the high-dimensional data given by the sequential and multi-class records of each patient. Each of the patient case series input to the classifier is of different length and different diagnosis times. Standard classifiers cannot deal with these multi-value measurements, and classifiers cannot deal with multiple classes for each patient. This research aim to make the classification results dependent on only a single time measurement. That is, each measurement from each patient is mapped to a single value by applying a feature transformation method. The feature transform can be computed to map a new set of features from the original raw data. This new set is the representation of data provided to the classification task, and will
increase the accuracy of the classifier.

Thus, in this paper, our main contribution is a new feature values that represents the temporal data under the inner distance combination transform (IDCT) method. This concept uses time sequence data to form a single value before classification. The efficiency of the approach is compared with mean, median, mode, final time treatments for the transform method and the final set of values in an obesity dataset. Finally, a naïve Bayes classifier is trained using the new feature set.

This paper proceeds as follows. Section 2 summarizes the problem and related work, and Section 3 introduces the novel IDCT approach for mining temporal medical data for classification. Section 4 examines the evaluation accuracy using temporal obesity data, and Section 5 concludes our work.

2. Related Works

This section summarizes various reviews and technical on temporal medical process and data mining techniques.

Ernst, H., and Gert, P., [7] described the classification of electroencephalograph (EEG) signals for brain–computer interfaces (BCIs). Their paper compared two different neural network topologies for classifying single EEG data from a BCI. They compared standard multilayer perceptron classifiers using a finite impulse response (FIR) network of single-trial EEG classification, and showed that the error rates and robustness of the FIR classifiers improved the performance.

Hirano, S., et al. [9] described a method for analyzing time-series laboratory examination databases. In experiments, show that hybrid approach can be used to discover interesting temporal patterns hidden in time-series databases. The key concept of their method is to classify temporal patterns using multi-scale structure matching and a rough set-based clustering method. Multi-scale matching enables the similarity between two sequences of examinations to be captured from both short-term and long-term perspectives. The rough set-based clustering technique was then applied to classify the sequences according to the relative similarity obtained through multi-scale matching.

Peter, R., et al. [19] described a new temporal data classification method for phenomena that exhibit major trends that develop gradually over time and contain significant fluctuations in the measured values at adjacent time instances. Weather is a good example of such a temporal phenomenon because of clear warming or cooling trends over weeks or months occurring simultaneously with significant daily fluctuations. The experiments used the Texas Commission on Environmental Quality (TCEQ) database, which recorded meteorological data between 1998 and 2004. Experiments on the TCEQ database showed two major results: (1) significant accuracy improvements were obtained using historical data; and (2) no accuracy improvement was obtained by the use of more than three features.

Bellazzi, et al. [3] This paper review the main approaches to mine biomedical time sequences described how temporal data mining is becoming an important tool for healthcare providers and decision makers. The ability to handle and analyze complex multivariate data may allow useful information to be extracted from day-to-day healthcare activities as well as from patient monitoring. This paper review the main approaches to mine biomedical time sequences.

Batal and Valizadegan [12] proposed a pattern-based classification framework for multivariate time series data. Their approach relies on temporal abstractions and temporal logic to construct the classification features, and uses the minimal predictive temporal pattern framework to present an efficient algorithm for directly mining these patterns.

3. Methodology

For the problem of temporal classification, a sequence of input values is nonlinearly expanded to a higher dimension and mapped to a desired output sequence. Classifier maps a feature set of patient case series to a set of class labels. The structure of a patient case series containing temporal medical data is a sequence that is typically recorded chronologically at fixed intervals. The methodology proposed a novel inner distance combination transform method (IDCT). The method is divided into two steps: the feature transform and classification. First, the feature values take time sequence data generated in the patient subspace and forms a new values in which one or all singular values are fixed by an inter distance norm combination transform method. Second, a naïve Bayes classification is performed based on the novel extracted features. This process identifies the class
structure of a temporal sequence by considering the feature values varying in time, which improves the overall accuracy [8]. Figure 1 illustrates the classification process for patient case data.

Figure 1. Block diagram illustrating the proposed method.

3.1. Inner Distance Combination Transform Method (IDCT)

In this section, we present the IDCT method for constructing a representative feature value. We transform the subspace patient case series $PID-X_i$ to define a mapping from the input feature subspace into a new feature values $X_i'$. $X_i'$ is used instead of $X_i[t_j], j = 1..n$, for training the classifier. The new feature values given by the transform method are fed into a classifier model. This provides more accurate predictions of the class labels for future patients than the original raw data, which can be used to enhance the classification performance. Figure 2 shows the merge level of the data under an inner distance function, and the combination of each set in the series for the transform.

$$X_i[t_j] = \{ X_i[t_1] X_i[t_2] \ldots X_i[t_{n/2-1}] X_i[t_{n/2}] \} \cdot \{ X_i[t_{n/2+1}] X_i[t_{n/2+2}] \ldots X_i[t_n] \}$$

Figure 2. Merging combined with the inner distance function.
IDCT consists of computational analyses with an inner distance function, and combines a set of series of the real values. Given PID-0n- $X_i[t_j]$ is a set of patient case series, according to the number of time treatment or subspaces for classified into classes, each having measurement ($X_i$) and a time-visit of patient $t_j$ in each measurement of patient case series, Suppose each series has a Time point($T_j$), $T_j = \{t_1, t_2, t_3, ..., t_n\}$, where $j$ is the visit of a patient and $n$ is the total number of visits. Each patient case series covers a different length of time. Thus, we map the raw input data to generate a new representation. In the case of continuous inputs $X_i$ and a dataset involving feature time sequences, the original data set is transformed according to the newly created features and then used for mining. $X_i \rightarrow X_i'$ maps each sequence $X_i$ to a fixed feature values $X_i'$. Suppose $a_{iti}$ and $b_{iti}$ are set of values in each measurement, $D$ is the result of combining the results of the $A$ and $B$, represented by $X_i'$ from one patient case series. The output of IDCT is a new feature value representing by $X_i'$ from all patient case series. This is easy to interpret, and can be trained to predict the patient class.

These procedures are as follows.

---

**Procedure:** Inner distance combination transformation method

**Input:** a complete temporal dataset ($D$)

**Output:** singular new feature values

---

**Step 1.** Separate data in the patient subspace matrix.

**Step 2.** Repeat

**Step 3.** Select each patient in the subspace matrix.

3.1 Transpose the patient subspace matrix.

3.2 Repeat

3.2.1 $X_i\ ; \quad i = i+1, \ i=1,2,3,\ldots,m$

3.2.2 Select a row value in the subspace matrix.

$$[X_i[t_1] \ X_i[t_2] \ X_i[t_3] \ X_i[t_4] \ldots X_i[t_n]\ ] \ , \ i=1 \ldots m$$

3.2.3 Separate a set in the series into two sets.

$$[X_i[t_1] \ X_i[t_2] \ X_i[t_3] ... X_i[t_{n/2}]\] \quad \text{and} \quad [X_i[t_{n/2+1}] \ X_i[t_{n/2+2}] \ X_i[t_{n/2+3}] ... \ X_i[t_n]\]$$

3.2.4 Compute $A$ and $B$ from the first set in 3.4

In each level $j$; $j = 1,2,3,\ldots,n/2$.

\[
\begin{align*}
    a_{[1,k]} &= \begin{cases} 
        \sqrt{x_i[t_k]^2+x_i[t_{k+1}]^2} , & k = 1,2,3,\ldots,n/2-1 \\
        x_i[t_k]^2+x_i[t_{1}]^2 , & k = n/2 
    \end{cases} \quad \ldots \ldots (1) \\
    b_{[1,k]} &= \begin{cases} 
        \sqrt{x_i[t_k]^2+x_i[t_{k+1}]^2} , & k = n/2+1, n/2+2,\ldots,n-1 \\
        x_i[t_k]^2+x_i[t_{n/2+1}]^2 , & k = n 
    \end{cases} \quad \ldots \ldots (2)
\end{align*}
\]

when $a_{[j,k]}$ and $b_{[j,k]}$ where $j = 2,3,4,\ldots,n/2$

\[
\begin{align*}
    a_{[j,k]} &= \begin{cases} 
        a_{[j,k]}^2+a_{[j,k+1]}^2 , & k = 1,2,3,\ldots,n/2 \\
        a_{[j,k]}^2+a_{[j,1]}^2 , & k = n/2 
    \end{cases} \quad \ldots \ldots (3) \\
    b_{[j,k]} &= \begin{cases} 
        b_{[j,k]}^2+b_{[j,k+1]}^2 , & k = 1,2,3,\ldots,n/2 \\
        b_{[j,k]}^2+b_{[j,1]}^2 , & k = n/2 
    \end{cases} \quad \ldots \ldots (4)
\end{align*}
\]
3.2.5 Compute A and B

\[ ||A|| = \text{Max} (a_{i[j,k]}) , \ j = n/2 , \ k = 1,2,3,\ldots,n/2 \]  
\[ ||B|| = \text{Max} (b_{i[j,k]}) , \ j = n/2 , \ k = 1,2,3,\ldots,n/2 \]  

3.2.6 Compute D

\[ D(X_i) = \text{Max} (||A||, ||B||) \]  

3.2.7 Until \( i < m \).

Step4. \( X_i = [(x_1') \ (x_2') \ldots \ (x_n')] \) //new feature

Step5. Repeat step 2 for patient subspace \( n+1 \).

Step6. \( X'_i = [(x_1') \ (x_2') \ldots \ (x_n')] \) //new feature set

3.2. Naive Bayes classification

The naïve Bayes (NB) algorithm is based on conditional probabilities. The NB classifier is one of the most effective and efficient classification algorithms, and has been successfully applied to many medical problems. It is a probabilistic classifier based on Bayes’ theorem, which calculates a probability by counting the frequency of values and combinations of values in the historical data.

**Theorem:**
Bayes’ Theorem finds the probability of an event occurring given the probability of another event that has already occurred. Let \( C \) be a conditional probability, the likelihood of some conclusion, and \( E \) be some given evidence/observation. We assume there is a dependence relationship between \( C \) and \( E \).

This probability is denoted as \( P(C | E) \), where

\[ P(C | E) = \frac{P(E | C) P(C)}{P(E)} \]

**NB classifier:** Suppose that there are \( m \) classes \( C_1, C_2, \ldots, C_m \). Given a tuple \( X \), the classifier will predict that \( X \) belongs to the class having the highest posterior probability, conditioned on \( X \). That is, the NB classifier predicts that tuple \( X \) belongs to class \( C_i \):

\[ P(C_i|X) > P(C_j|X), \text{ for } 1 \leq j \leq m, j \neq i \]

By Bayes’ Theorem:

\[ P(C_i | X) = \frac{P(X | C_i) P(C_i)}{P(X)} \]

and under conditional independence:

\[ P(X | C_i) = \prod_{k=1}^{n} P(x_k | C_i) \]

Thus, we have the probability:

\[ = P(x_1 | C_i) \times P(x_2 | C_i) \times \ldots \times P(x_n | C_i) \]

With an NB classifier, we can further estimate the probability. To predict the class label of \( X \), \( P(X|C_i) P(C_i) \) is evaluated for each class \( C_i \). The classifier predicts that the class label of tuple \( X \) is the class \( C_i \):

\[ P(X|C_i) P(C_i) > P(X|C_j) P(C_j), \text{ for } 1 \leq j \leq m, j \neq i \]

The predicted class label is the class \( C_i \) for which \( P(X|C_i) P(C_i) \) is the maximum. NB uses the argmax function to calculate the maximized probability, called the maximum posteriori hypothesis, from both the classes. The corresponding classifier is the function that classifies every new instance defined as follows \([1, 2, 5, 15, 18]\):

\[ V_{NB} = \text{argmax}_{K} P(C_i) \prod_{k=1}^{K} P(x_k | C_i) \]
4. Experiments

We conducted the following experiments. First, we used a complete dataset and applied the transform method to give a singular value for the classifier. This transformation increased the classification accuracy. Second, we compared the prediction performance of our IDCT method with five well-known methods using a complete obesity dataset.

4.1. Dataset

In this research, we used a dataset containing 450 cases and approximately 1,200 records from patients with obesity at the Cardiovascular and Metabolism Center, Ramathibodi Hospital. The data consist of re-code numbers, patient IDs, and treatment dates, as well as the patient’s age, sex, height, weight, body mass index (BMI), basal metabolic rate, skeletal muscle mass, percentage of body fat, waist/hip ratio, edema examination, target control, weight control, fat control, muscle control, fitness score, body mass, and protein (g).

Table 1. Structure of obesity medical temporal data

<table>
<thead>
<tr>
<th>PID</th>
<th>Time</th>
<th>Weight</th>
<th>BMI</th>
<th>…</th>
<th>…</th>
<th>Xₙ</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>…</td>
<td>…</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>x</td>
<td>x</td>
<td>-</td>
<td>…</td>
<td>…</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>x</td>
<td>x</td>
<td>-</td>
<td>…</td>
<td>…</td>
<td>1</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>x</td>
<td>-</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>0</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>n</td>
<td></td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>

A set of patient visits is defined as \( T_j = \{t_1, t_2, t_3, \ldots, t_n\} \), where \( t_i \) is the \( i \)-th visit of patient \( j \) and \( n \) is the total number of visits. Personal patient data are collected and arranged in an \( m \times n \) matrix. In the time series \( X = \{x_1, x_2, \ldots, x_n\} \), \( X \) is the measurement variable, \( x_i \) is the recorded value of measurement variable \( X \) at time \( i \), and \( n \) is the number of observations. Each event occurring at each time point has a recorded value. Formally, a patient case series in temporal medical data can be represented by \( \text{PIDxx-X} = \{x_{i1}, x_{i2}, \ldots, x_{in}\} \).

4.2. Classifier evaluation

4.2.1. Performance measure

Studies on the classification accuracy of temporal medical data often yield continuous data based on class predictions. Accuracy is the most important criterion in the classification literature, because the main goal is to correctly classify an unknown instance of temporal data. The mined features dramatically improve the performance accuracy of the classifiers.

The performance of the NB classifier is measured using the obesity dataset. The temporal medical data are stratified ten-fold cross validation for estimating accuracy. We evaluate the accuracy of the model to measure the performance of the classifiers. The accuracy rate is measured on the transformed data set. The classifier efficiency is computed with regard to the true or false classification results [8], with some common metrics defined using the confusion matrix shown in Table 2.

Table 2. Confusion matrix

<table>
<thead>
<tr>
<th>Predict class</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>True Positive (TP)</td>
<td>False Positives (FP)</td>
</tr>
<tr>
<td>Negative</td>
<td>False Negative (FN)</td>
<td>True Negatives (TN)</td>
</tr>
</tbody>
</table>
1. The precision of a classifier is the percentage of positive predictions made by the classifier that are correct.

\[ \text{Precision} = \frac{TP}{TP+FP} \]

2. Recall is the percentage of true positives that are correctly detected by the classifier.

\[ \text{Recall} = \frac{TP}{TP+FN} \]

3. F-measure is defined as the harmonic mean of the recall and precision:

\[ \text{F-measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \]

4. Accuracy measures the ability of the classifier to correctly classify unlabeled data:

\[ \text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \times 100\% \]

4.2.2. Experimental Results

New feature values represent temporal data for the NB classifier. We compare the proposed IDCT method with the mean, median, mode, and the final record of each patient (Tn). The structure of results for our transform method is given in Table 3.

<table>
<thead>
<tr>
<th>PID</th>
<th>IDCT</th>
<th>Mean</th>
<th>Fₐ</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X₁'</td>
<td></td>
<td>Xₙ'</td>
<td></td>
</tr>
<tr>
<td>01</td>
<td>..</td>
<td>..</td>
<td>..</td>
<td>..</td>
</tr>
<tr>
<td>02</td>
<td>.</td>
<td>..</td>
<td>..</td>
<td>..</td>
</tr>
<tr>
<td>...</td>
<td>..</td>
<td>..</td>
<td>..</td>
<td>..</td>
</tr>
<tr>
<td>N</td>
<td>..</td>
<td>..</td>
<td>..</td>
<td>..</td>
</tr>
</tbody>
</table>

Table 3. Structure of transform method results

<table>
<thead>
<tr>
<th>Method</th>
<th>NB (Naive Bayes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-Inner distance combination-IDCT (Propose)</td>
<td>94.435 0.945</td>
</tr>
<tr>
<td>-Mean</td>
<td>70.613 0.716</td>
</tr>
<tr>
<td>-Median</td>
<td>80.782 0.804</td>
</tr>
<tr>
<td>-Mode</td>
<td>82.805 0.829</td>
</tr>
<tr>
<td>-Final (Tₙ)</td>
<td>86.305 0.864</td>
</tr>
</tbody>
</table>

Table 4. Accuracy of our results

The NB classifier is generated by the first set of records containing treatment information for patients who reach the last treatment. The experimental results show that the transformation technique increased the classification accuracy. The accuracy of IDCT is 94.435%, which is better than that of the other methods shown in Table 4.

5. Conclusion

In this work, we have presented a novel transform method based on a feature value for the prediction of temporal classes. The aim of this research was to provide a new and complementary set of features for the current problem. The proposed approach is divided into two different tasks: the feature transform, and classification. The feature transformation technique generates an inner distance
combination of obesity values. This allows us to construct a new feature set of singular values to enable better classification accuracy. Experiments were performed using temporal obesity data. The results of a confusion matrix test showed that the proposed IDCT method achieves better accuracy than existing methods. The proposed approach can also be applied for classification.

6. Acknowledgment

This research was supported by the Higher Education Commission scholarship. The data was from the Cardiovascular and Metabolism Center, Ramathibodi Hospital.

7. References


