Case based time series prediction using biased time warp distance for electrical evoked potential forecasting in visual prostheses

Jin Qi\textsuperscript{a,}\textsuperscript{*}, Jie Hu\textsuperscript{a}, Yinhong Peng\textsuperscript{a}, Xinyu Chai\textsuperscript{b}, Qiushi Ren\textsuperscript{c}

\textsuperscript{a} Institute of Knowledge Based Engineering, School of Mechanical Engineering, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China
\textsuperscript{b} Institute of Laser Medicine and Bio-Photonics, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China
\textsuperscript{c} Department of Biomedical Engineering, Peking University, 60 Yannan Yuan, Beijing 100871, China

\textbf{A R T I C L E   I N F O}

\textbf{Article history:}
Received 20 October 2011
Received in revised form 8 October 2012
Accepted 25 November 2012
Available online 11 December 2012

\textbf{Keywords:}
Time series prediction
Electrical evoked potential
Temporal similarity measure
Distance metric
Bias configuration

\textbf{A B S T R A C T}

Case based time series prediction (CTSP) is a machine learning technique to predict the future behavior of the current time series by referring similar old cases. To reduce the cost of the visual prostheses research, we devote to the investigation of predictive performance of CTSP in electrical evoked potential (EEP) prediction instead of doing numerous biological experiments. The heart of CTSP for EEP prediction is a similarity measure of training case for target electrical stimulus by using distance metric. As EEP experimental case consists of the stationary electrical stimulation values and time-varying EEP elicited values, this paper proposes a new distance metric which takes the advantage of point-to-point distance’s efficient operation in stationary data and time series distance’s high capability in temporal data, called as biased time warp distance (BTWD). In BTWD metric, stimulation set difference (DiffJ) and EEP sequence difference (DiffJ) are calculated respectively, and a time-dependent bias configuration is added to reflect the different influences of DiffJ and DiffJ to the numerical computation of BTWD. Similarity-related adaptation coefficient summation is employed to yield the predictive EEP values at given time point in principle of $k$ nearest neighbors. The proposed predictor using BTWD was empirically tested with data collected from the electrophysiological EEP eliciting experiments. We statistically validated our results by comparing them with other predictor using classical point-to-point distances and time series distances. The empirical results indicated that our proposed method produces superior performance in EEP prediction in terms of predictive accuracy and computational complexity.

\textcopyright 2012 Elsevier B.V. All rights reserved.

1. Introduction

1.1. The feasibility of applying case based time series prediction (CTSP) in visual prostheses

As the interests in time series have been explored by the number of domains in which time series occur: medicine, industry, entertainment, finance, computer science, meteorology and in almost every other field of human activity [1], the problem of time series prediction (TSP) has attracted great attention [2–5], in which the historical observation values are collected and analyzed to predict the possible values of future time point. Different from traditional TSP approaches based on neural network (NN) model or regression model, case based TSP (CTSP) allows us to use previous data in predicting current data without completely understanding the underlying mechanism of time series behavior [6], actually, it would be very useful for experts to make their decision and predict the future behavior of the current time series by referring similar old case from past.

The purpose of optic nerve visual prosthesis is to deliver the electrical impulses to optic nerve with the penetrating electrode, and the visual perception is then emerged from the visual cortex, therefore, the analysis of responses of the visual cortex after the electrical stimulation to the optic nerve is a key part in optic nerve visual prosthesis [7–10]. To explore the temporal property of responses, electrical evoked potentials (EEPs) are elicited over the rabbit skull when the optic nerve was stimulated with specific electrical stimulation current, as shown in Fig. 1. However, subjected to experiment cost and material restriction, the EEP response data triggered by various stimulations are too sparse and inadequate to be analyzed, and the characteristics of EEP data are time series, non-linear, inherently noisy, non-stationary, and deterministically chaotic [11]. To reduce the cost of the visual prostheses research, it is useful to exploit existing EEP electrophysiological data to predict new EEP response at given time point under new electrical stimulations, e.g., the current intensity, the pulse duration, the frequency...
of the stimuli, etc. In our previous works [11,12], we presented an adapted support vector machine (SVM) based TSP model to predict EEP movement but with higher computational complexity. Enlighten by the advantages of CTSP, we devote to the investigation of predictive performance of CTSP to solve the EEP forecasting problem.

1.2. The contribution of this research

In this research, we attempt to employ CTSP to undertake EEP prediction task, and each experimental sample is expressed as an EEP experimental case, which involves stationary electrical stimulation values and time-varying EEP elicited values. The contribution of this paper is twofold. On the one hand, we try to construct a new distance function of similarity measure for EEP prediction. For the sake of implementing CTSP, training case should capture the evolution of the observed phenomenon over time [13], but time-varying information in cases will be neglected or the retrieval procedure will require massive calculated quantity when classical point-to-point distance or time series distance is used to measure the similarity of each case. The motivation of this research lies in employing a novel distance metric to calculate the similarity of EEP biological case by referring classic distance metrics, and intending to introduce the bias parameter to reflect the different influences from stationary electrical stimulation values and time series EEP elicited values for new predictive result, so the proposed distance is called biased time warp distance (BTWD). This paper is the first attempt to develop a new similarity measurement in CTSP which takes the advantages of classical point-to-point distance function's efficient operation in stationary data and time series distance function's high capability in temporal data.

On the other hand, we try to investigate the superiority of BTWD in actual EEP prediction, and the predictive performances of CTSP with BTWD and other TSP approaches are planned to be compared. Initial EEP experimental data are collected from the studies of rabbit’s visual cortex by optic nerve stimulation carried out inside the pia mater, on the pia mater and on the dura mater. In the assessment of predictive performance, we employ 30-time cross-validation strategy by combining hold-out method and leave-one-out cross-validation. Predictive performance is evaluated by predictive results produced on hold-out data. Statistical analysis is employed to find whether or not there are significant differences among comparative predictors in the light of the predictive results.

The breakdown of this paper is organized as follows. The next section summarizes the research trends in TSP, and the motivation and originality of this research. Section 3 presents the specification on the BTWD. Section 4 describes the empirical design. Section 5 presents the empirical results and relevant discussions. The last section provides the conclusions and suggestions for further research.

2. Related works

2.1. Model based time series prediction

Model based TSP (MTSP) represents a family of TSP methods based on various predictive models. Initially, TSP for business failure prediction has been extensively studied from the views of statistic models in 1960s [14], e.g., discriminant analysis model [15,16] and logistic regression model [17] were applied to predict business failure. Later, moving average model [18] and auto-regressive integrated moving average model [19] have been proposed to forecast the monthly electricity demand, but statistic models only work well when the data dependency is linear [11]. To coincide with non-linear fluctuation path of time series data, the neural network (NN) models have been developed for time series forecasting since the 1990s [20–25], which apply the empirical risk and mean squared error minimization in some way. Chang et al. [2,26,27] argued that the multi-layer perceptron NN model is the most commonly used NN model for TSP. Although NNs work well, they have some inherent drawbacks, for example the problem of multiple local minima, the choice of number of hidden units and the danger of overfitting [28–30].

Another widely used non-linear model for TSP is support vector machine (SVM) model [31,32]. Unlike NN model, SVM implements the structural risk minimization and tries to minimize an upper bound of generalization error instead of minimizing the misclassification error or deviation from correct solution of the training data. Thus, SVM could achieve an optimum network structure, and resulting in better generalization performance than NNs [33–36]. Under this principle, the SVM based TSP for EEP prediction has been presented in our previous works [11,12], but we found that the over-complex SVM model would be computationally intensive. Moreover, MTSP operates training data implicitly in black box and remains difficult to explain the final predictive results.

2.2. Case based time series prediction

Compare with MTSP approach, CTSP is believed to be a preeminent method for predicting temporal activity by the following reasons [37]. First CTSP is considered as a non-parametric method which dose not requires any time-stamp data distribution assumption for input temporal case. Second, CTSP is an incremental learning technique that can retain new case without reprocessing to update the previous case base. By contrast, MTSP methods like SVM are batch-oriented, wherein both new and old time series data must be submitted as a single batch to the model. And third, CTSP not only can process data more explicitly in explaining analytic results but also can address better efficiency in predicting dynamic fluctuant problem [38].

Performance of CTSP depends largely on the mechanism of similarity measure. Classical point-to-point distance functions such as Euclidean distance (ED) [28,39–41] and Manhattan distance.
(MD) [42] have been employed to measure the similarity in the process of case retrieval of CTSP, where the temporal sequence of case is viewed as a series of points in an appropriate multi-dimensional space. Other distance measures used in time series clustering, as reviewed by Liao [43], could also be used here. However, time-varying information embedded in temporal cases has been neglected and its role also has been oversimplified in traditional CTSP. To handle the time series information embedded in training case, temporal dimension reduction approaches such as discrete Fourier transform [44,45], discrete wavelet transform [46], singular value decomposition [47], piecewise linear approximation [48] and symbolic aggregate approximation [49,50] are employed to transfer the two-dimensional time series sequence into one dimension space which consists of a series of points, and then the classical ED metric is utilized to compute the similarity.

Later, to achieve a better measurement effect, time series distance metrics have been developed to calculate the similarities among time series sequences without temporal dimension reduction, such as dynamic time warping (DTW) distance [49–55], cosine distance [56–58] and Lp norm [59–61]. Bouchenah [62] designed a shape exchange algorithm (SEA) in distance function, which is able to perform a point-to-point matching and render a kind of difference between the two time series. However, Lp norm and SEA have the same weakness as the ED metric in that it difficult to capture the fluctuation trends in time series data [61]. Among the time series distances, DTW is the most popular way, but it dose not satisfy the triangular inequality which is a desirable property to speed up retrieval procedure [61,62].

Another category of time series distance function appropriates for time-stamp data is edit distance, developed by Levensstein firstly in 1966 [63]. Unlike other time series distances, this distance is defined as the minimum number of edit operations needed to transform a sequence into another [61,64,65]. Following that, the extension methods of edit distance have been proposed, e.g., longest common sub-sequence (LCSS) [66], edit distance with real penalty (ERP) [67], and point-patterns matching (PPM) [68]. Recently, Marteau [60] offered a time wrap edit distance (TWED) method integrating with DTW, Lp-norm and ERP. TWED addresses the similarity between two time series sequences is measured as the minimum cost sequence of edit operations. So far, these edit distance extension approaches mainly pay attention to temporal dimension, but not to the multi-dimensional space.

To support electrical stimulation space and temporal EEP elicited space simultaneously, in this study, we concentrate on introducing a new distance function through referring ED and TWED metrics to figure out the similarity between two cases. Moreover, we intend to introduce the bias parameter to reflect the different influences from stationary electrical stimulation values and time series EEP elicited values for new predictive result, so the proposed distance is called biased time warp distance (BTWD). Finally, we also attempt to provide empirical evidence on the performance of BTWD for EEP forecasting, and the experimental data are collected to compare the predictive performances with other CTSP methods using traditional distances.

3. Case-based time series prediction using biased time wrap distance

3.1. Framework of CTSP using BTWD for EEP prediction

The framework of CTSP process using BTWD metric is shown in Fig. 2, where initial cases collected from EEP electrophysiological recording data is constructed, containing several stimulation values and EEP values. Accordingly, there are two types of differences in case comparison problem, that is, stimulation set difference (Diff_I) and EEP sequence difference (Diff_Ii). Diff_I means the difference of electrical stimulation sets from two comparison cases, and Diff_Ii is the difference between two temporal EEP elicited sequences of cases. Thus, the general BTWD value between two historical cases is composed of Diff_I and Diff_Ii. Meanwhile, the bias parameter is added to emphasize the different influences of Diff_I and Diff_Ii to the numerical computation of BTWD, whose value will change over time. So a time-dependent bias configuration should be also considered in this paper. Finally, the predictive EEP values for target stimulus input can be generated by similarity-related summation on the foundation of k similar cases.

![Fig. 2. Framework of CTSP using BTWD metric for EEP prediction, which consists of three major elements, i.e., case representation, case retrieval and case reuse.](image-url)
3.2. Case retrieval using biased time warp distance

3.2.1. Mechanism of biased time warp distance

The heart of CTSP is the similar case searching which guides future prediction and problem solving. The concept of similarity between two cases explained in this way [14]: the less difference between two cases, the more similar they are. In this research, we present the BTWD metric to embody the difference of two EEG experimental cases. Given two EEG experiment cases \( \tilde{A} = (\tilde{A} \oplus \tilde{A}) \) and \( \tilde{B} = (\tilde{B} \oplus \tilde{B}) \), which contain stimulation trigger parts, i.e., \( \tilde{A} \) and \( \tilde{B} \), and EEG elicited response parts, i.e., \( \tilde{A} \) and \( \tilde{B} \). Note that, \( \tilde{A}, \tilde{B} \in S \times T \) and \( S \subset \mathbb{R}^n \). To calculate the BTWD value between \( \tilde{A} \) and \( \tilde{B} \) expressed as \( d(\tilde{A}, \tilde{B}) \), can be calculated as follows:

\[
d(\tilde{A}, \tilde{B}) = |1 - \alpha(t)| \times \text{diff}'(\tilde{A}, \tilde{B}) + \alpha(t) \times \text{diff}''(\tilde{A}, \tilde{B})
\]  

(1)

where \( \text{diff}'(\tilde{A}, \tilde{B}) \) and \( \text{diff}''(\tilde{A}, \tilde{B}) \) denote the DiffJ and DiffII between \( \tilde{A} \) and \( \tilde{B} \), and \( \alpha(t) \) is a bias configuration function to reflect the different influences of DiffJ and DiffII for computational result of \( d(\tilde{A}, \tilde{B}) \).

To calculate the DiffJ, the point-to-point distance function is employed as follows:

\[
diff'(\tilde{A}, \tilde{B}) = \left( \sum_{i=1}^{m} w_i \times |\tilde{a}_i - \tilde{b}_i|^2 \right)^{1/2}
\]  

(2)

where \( \delta \) is a distance adapted coefficient, for \( \delta = 2 \) we have the Euclidian one, while \( \delta = 1 \) is the Manhattan distance and \( \delta = \infty \) is the Chebychev distance, i.e., \( \text{Max}(|\tilde{a}_i - \tilde{b}_i|) \). \( \tilde{a}_i \) and \( \tilde{b}_i \) represent the values of \( i \)-th stimulus feature of \( \tilde{A} \) and \( \tilde{B} \) respectively, and \( w_i \) is the associated weight to this feature.

For EEG sequence difference denoted as DiffJ, we intend to refer the principle of TWED approach for a couple of reasons [60]: firstly TWED expresses a superior performance in dealing with local time shifting, secondly it satisfies the triangle inequality, an arithmetic property which might be valuable in computing DiffJ, finally the effectiveness of TWED has been empirically proved in [60,70,71]. More detailed descriptions on TWED could be obtained from [60,71]. To solve the problem of \( \text{diff}''(\tilde{A}, \tilde{B}) \) calculation, three operations, i.e., delete \( \tilde{A} \), delete \( \tilde{B} \), and match, are introduced, and \( \text{diff}''(\tilde{A}, \tilde{B}) \) is measured as the minimum cost sequence of editing operations needed to transform one time series \( \tilde{A} \) into \( \tilde{B} \). Thus cost function \( \Theta \) and cost minimum operation is added into \( \text{diff}''(\tilde{A}, \tilde{B}) \).

\[
\tilde{A} = (\tilde{a}_j)_{j=1}^{n} = ((av_j, t_j))_{j=1}^{n}, \quad \tilde{B} = (\tilde{b}_j)_{j=1}^{n} = ((bv_j, t_j))_{j=1}^{n}
\]  

(3)

where \( \tilde{a}_j \) is the \( j \)-th value-time pair of \( \tilde{A} \), that is, the EEG value \( av_j \) elicited at \( j \)-th time point \( t_j \) in case \( A \), and there are similar definitions for \( \tilde{b}_j \) and \( bv_j \) in case \( B \). \( n \) is the total number of time series value. In EEG biological experiments, the lengths of \( \tilde{A} \) and \( \tilde{B} \) are same. Hence, the difference between \( \tilde{A} \) and \( \tilde{B} \) with the same length \( n \) can be calculated by the following formula.

\[
diff''(\tilde{A}, \tilde{B}) = \text{diff}''(\tilde{A}^0, \tilde{B}^0)
\]

(4)

In formula (4), \( \Theta \) is the empty time series set. \( \tilde{a}_j \rightarrow \tilde{A}, \tilde{a}_j \rightarrow \tilde{B} \) and \( \tilde{b}_j \rightarrow \tilde{A} \) represent the operations of delete \( \tilde{A} \), match, and delete \( \tilde{B} \). The recursion of \( \text{diff}''(\tilde{A}, \tilde{B}) \) is initialized by \( \text{diff}''(\tilde{A}^0, \tilde{B}^0) = 0 \). Moreover, to facilitate the calculation, we also use ED metric to construct the cost function \( \Theta \), instead of \( l_p \)-norm metric used in TWED, and then the mathematic formula of \( \Theta \) can be expressed as follow:

\[
\Theta(\tilde{a}_j \rightarrow \tilde{A}) = \sqrt{(av_j - av_{j-1})^2 + \mu(t_j - t_{j-1})^2}
\]

(5)

\[
\Theta(\tilde{b}_j \rightarrow \tilde{A}) = \sqrt{(bv_j - bv_{j-1})^2 + \mu(t_j - t_{j-1})^2}
\]

\[
\Theta(\tilde{b}_j \rightarrow \tilde{B}) = \sqrt{(bv_j - bv_{j-1})^2 + \mu(t_j - t_{j-1})^2}
\]

\[
\Theta(\tilde{a}_j \rightarrow \tilde{B}) = \sqrt{(av_j - av_{j-1})^2 + \mu(t_j - t_{j-1})^2}
\]

(5)

where \( \mu \) is a non-negative constant penalty parameter, used to be proportional to the \( (t_j - t_{j-1}) \) measure during the edit procedure.

3.2.2. Time-dependent bias configuration function

According to previous studies, the movements of EEPs are not random, they behave in a highly non-linear and dynamic manner [79,11]. With further observation, we can see that the movements of EEG elicited curves are comparatively similar at the initial stage with specific stimulation setting, e.g., different pulse amplitudes but fixed pulse duration or different pulse durations but fixed pulse amplitude. As the elapse of time, the dissimilarities among EEG curves expand increasingly, as shown in Fig. 3. This means that the DiffII plays an increasing role over time in constructing the time wrapping distance. Conversely, the influence of DiffJ begins to decrease as time passed. To illustrate the time-varying influences of DiffJ and DiffII, we develop a time-dependent bias configuration function to determine the value of bias parameter, and the Gaussian function [39] is utilized to build the mathematical formula, expressed in the following way.

\[
\alpha(t) = \begin{cases} 
\exp(-\lambda(t-p)^2 + \ln(\omega)) & t \leq p \\
\omega & t > p \end{cases}
\]  

(6)

where \( \lambda \) is a parameter to control the ascending rate of \( \alpha(t) \) and \( \lambda > 0, p \) is a flexure time point, after that, the trend of \( \alpha(t) \) movement will level off, and the addition of natural logarithm \( \ln(\omega) \) also means the value of bias for DiffII will continue to increase before reaching a certain upper bound value \( \omega \in (0, 1) \), and the trend of bias value for DiffJ is just the opposite. Fig. 4 illustrates the \( \alpha(t) \) curves for DiffJ and DiffII, when \( p = 50 \) ms, \( \lambda = 0.01, \lambda = 0.04, \lambda = 0.002 \) and \( \lambda = 0.0 \). Due to the introduction of bias function, \( d(\tilde{A}, \tilde{B}) \) is also a time-dependent function, and its value at point \( t_j \) can be expressed as:

\[
d(\tilde{A}, \tilde{B}, t_j) = [1 - \alpha(t_j)] \times \text{diff}'(\tilde{A}, \tilde{B}) + \alpha(t_j) \times \text{diff}''(\tilde{A}^t, \tilde{B}^t)
\]

(7)

3.2.3. Similarity measurement based on biased time warp distance

After the value of \( d(\tilde{A}, \tilde{B}, t_j) \) has been computed, the similarity between case \( A \) and case \( B \) can be determined subsequently. Suppose that \( sim(\tilde{A}, \tilde{B}, t_j) \) is the similarity between two cases at time point \( t_j \), then the formulation of \( sim(\tilde{A}, \tilde{B}, t_j) \) can be expressed by the following formula.

\[
sim(\tilde{A}, \tilde{B}, t_j) = \frac{1}{1 + d(\tilde{A}, \tilde{B}, t_j)}
\]  

(8)
For EEP prediction, if case A and case B are defined as the target case and training case, then \( \text{sim}(A, B, t_j) \) also represents the similarity of case B for target case A at time point \( t_j \), which can be abbreviated as \( \text{sim}(B, t_j) \). The mechanism of similarity measure using BTWD is illustrated in Fig. 5.

### 3.3. Case reuse for electrical evoked potentials prediction

Similarity-related summation is employed to perform the case reuse process on the foundation of \( k \) similar cases. After the similarity of each training case in the case base has been calculated, a set of nearest neighbors can be retrieved [14]. The predicted EEP value is computed with a kind of similarity-related adaptation coefficient (AC) and then summed. The value of AC is directly proportional to the similarity between the old and target cases, and the sum of all ACs must be 1. The corresponding mathematical formula of AC under \( k \)-NN principle is calculated at first in the following way.

\[
AC(i)\big|_{t_j} = \frac{\text{sim}(case(i), t_j)}{\sum_{i=1}^{k} \text{sim}(case(i), t_j)}
\]  

(9)

where \( AC(i)|_{t_j} \) and \( \text{sim}(case(i), t_j) \) are AC value and similarity of case \( case(i) \) at time point \( t_j \).
Subsequently, the predicted EEP value at time point $t_j$, i.e., $EEP_{\text{prediction}}|t_j$, can be computed as:

$$EEP_{\text{prediction}}|t_j = \sum_{i=1}^{k} (AC(i)|t_j) \times (EEP_{\text{case}}(i)|t_j)$$

where $EEP_{\text{case}}(i)|t_j$ is the actual EEP elicited value of case(i) at time point $t_j$.

4. Empirical design

4.1. Objective and comparative methods

Objective of this empirical research is to investigate whether or not the proposed CTSP using BTWD metric can achieve higher EEP predictive performance compare with other CTSP methods using traditional distances. Two chief ways of distance computation in similarity measure of CTSP are point-to-point distance metrics and time series distance metrics. Thus, this study makes two empirical comparisons, i.e., comparison of predictive accuracy between predictors using BTWD and point-to-point distances, and analogous comparison between predictors using BTWD and time series distances. For the former comparison, classical point-to-point distance metrics such as ED and MD are adopted, and for the latter, DTW and TWED metrics are employed to perform the empirical comparison. Additionally, we utilize the distance name directly to represent the corresponding EEP predictor just for convenience. Hence, there are total five comparative EEP prediction methods in this empirical comparison, that is, BTWD, ED, MD, DTW and TWED. Meanwhile, we would like to investigate whether or not there is diversity in computational complexity between BTWD and other time series distance metrics including DTW, TWED, ERP and SEA, because the computational complexity is another important criterion for evaluating time series distance.

4.2. Data and variable

As we described above, each training case derived from EEP biological experimental involves trigger part and EEP response part, and we choose nine electrical stimulus features and EEP-time point pairs to express two parts respectively, as listed in Table 1. To guarantee the feasibility of this empirical comparison, totally 96 cases are collected, and each of which involves 500 time points starting from the earliest time point 0.1 ms with 0.1 ms time interval, and we use MATLAB toolkits to implement the proposed algorithm for carrying out the comparisons.

4.3. Parameter optimization of proposed method

According to the Formula (6), (7) and (9), there are five parameters should be optimized in the process of BTWD computing, that is, flexure point p, upper bound value $\omega$, constant value $\lambda$, constant penalty $\mu$, and nearest neighbor number k. Among them, $p$ and $\lambda$ determine the shape of $\alpha(t)$ curve, while $\omega$, $\mu$ and $k$ impact the efficiency of predictive performance. In this empirical comparison, 50 ms is the upper bound of selected time point, thereby it is a logical choice to use 50 ms as the value of $p$. Meanwhile, as indicated in Fig. 5, the reasonable range of $\lambda$ value is [0.002, 0.01], here the value of $\lambda$ is 0.004, and by reference to previous studies \cite{14,28,39,41}, we set $k=7$. For $\omega$ and $\mu$, it is not known beforehand which values of $\omega$ and $\mu$ are the best for one problem. This study prefers a grid-search on $(\omega, \mu)$ using leave-one-out cross-validation for model parameter selection. Each time one distinct sample is taken out to act as target case, while the rest consist of case base. The predictive EEP values generated on the foundation of case base are compared with its true EEP values from target case to determine the cross-validation accuracy rate (CVAR). Let the original values of $\omega$ and $\mu$ are (0.5, 0.55, . . . , 0.9) and (0.3, 0.35, . . . , 0.75) empirically and 500 time points are selected, then the grid-search is performed respectively in training set containing 66 experiment cases and test set containing 30 experiment cases to figure out the optimal solution. The CVAR of each $(\omega, \mu)$ option is calculated as follows:

$$CVAR = \frac{1}{n} \left( \sum_{k=1}^{n} \sum_{i=1}^{500} \frac{1}{500} (1 - ((a_{k,t} - p_{k,t})/a_{k,t})) \right)$$

for training set $n = 66$

for test set $n = 30$

where $a_{k,t}$ and $p_{k,t}$ represent the actual and predictive EEP value at $t$th time point of $k$th cross-validation, and $n$ is the number of cases.

If $(\omega, \mu)$ option lead to a maximal cross-validation rate estimated for the training set and test set, then it will be considered as the best solution. We selected six candidate $(\omega, \mu)$ pairs in order with highest cross-validation rates in training set and test set respectively, as shown in Fig. 6, and it is easy to see that (0.8, 0.55) and (0.85, 0.5) have higher cross-validation rates both in training set and test set, and we choose (0.85, 0.5) as the final solution in this example.

4.4. Validation technique and evaluation criteria

In this empirical example, we use the rabbit EEP experimental case with the most common values of electrical simulation as the target case, illustrated in Table 2. Its corresponding $F1$–$F9$ stimulation values are considered as the target stimulus parameters, and total 500 EEP temporal values of F10 are compared with the predictive results produced by comparative predictors on basis of the test set consists thirty cases. We define point predictive accuracy (PPA) and segment predictive accuracy (SPA) as evaluation criteria to evaluate the predictive performance. However, the conventional absolute percentage arithmetic of PPA ignores the evolution of the time series EEP value, so we will add a direction evaluation factor into PPA algorithm, besides single-point absolute percentage. In return, the mathematic expression of new PPA can be described
in the following formula.

\[
P_{PPA} = 1 - \left( \frac{1 + \psi_i^1}{2} \right) \times \frac{|a_{i-1} - p_{i-1}|}{a_{i-1}} + \frac{|a_i - p_i|}{a_i} + \left( \frac{1 + \psi_i^2}{2} \right) \times \frac{|a_{i+1} - p_{i+1}|}{a_{i+1}} \quad 1 < i < 500 \quad i \in N + \tag{12}
\]

And the PPA calculation is initialized as follows:

\[
P_{PPA} = 1 - \left( \frac{|a_{i} - p_{i}|}{a_{i}} \right) + \left( \frac{1 + \psi_i^1}{2} \right) \times \frac{|a_{i+2} - p_{i+2}|}{a_{i+1}} \]

\[
P_{PPA500} = 1 - \left( \frac{1 + \psi_{500}^2}{2} \right) \times \frac{|a_{499} - p_{499}|}{a_{499}} + \frac{|a_{500} - p_{500}|}{a_{500}} \tag{13}
\]

where \(a_i\) and \(p_i\) is the actual EEP value and predictive EEP value at \(i\)th time point, \(PPA_i\) is the corresponding PPA value at \(i\)th time point. \(\psi_i^1\) and \(\psi_i^2\) are lower error direction coefficient and upper error direction coefficient for predictive value \(p_i\), and their values are defined as follows:

\[
\psi_i^1 = \begin{cases} 
1 & (a_i - a_{i-1})(p_i - p_{i-1}) < 0 \\
0 & \text{others}
\end{cases} \tag{14}
\]

\[
\psi_i^2 = \begin{cases} 
1 & (a_{i+1} - a_i)(p_{i+1} - p_i) < 0 \\
0 & \text{others}
\end{cases} \quad 1 < i < 500 \quad i \in N + \quad \text{and} \quad \psi_0 = \psi_{500} = 0 \tag{14}
\]

After the definition of PPA has been determined, the formula for SPA criterion can be expressed as follows:

\[
SPA_{i,j} = \sum_{j=1}^{500} \frac{PPA_{i,j}}{j-i+1} \quad 1 \leq i \leq j \leq 500 \quad i,j \in N + \tag{15}
\]

where \(SPA_{i,j}\) is the predictive accuracy of time series segment \(i \rightarrow j\), which demonstrates the average PPA value from \(i\)th time point to \(j\)th time point, and \(j - i + 1\) is the length of segment \(i \rightarrow j\).

Moreover, to avoid the prejudices of one time hold-out method, 30 times hold-out method [11,12,14,71] is employed in test set to create 30 hold-out testing data to implement the comparison.
Considering the computational burden, we only pick out 50 time points from 1 ms to 50 ms with 1 ms interval, i.e., \( t_i = \{1 \text{ ms}, 2 \text{ ms}, \ldots, 50 \text{ ms}\} \). For each hold-out time, one case from test set is randomly selected as target case and other cases compose into the source cases. Thus, \( 30 \times 50 \) EEP predictive results will be achieved for each predictor. Furthermore, the Theil inequality coefficient (TIC) \([11,72,73]\), a statistical method, is employed to compare the results of two comparative predictors and to validate the predictors. Unlike the simple sum of squared errors, the principal advantage of the TIC is that it varies between 0 and 1, where values closer to zero indicate a better model validity \([74]\). The definition of TIC is displayed as follows:

\[
\text{TIC} = \frac{\sum_{i=1}^{50} (a_{ti} - p_{ti})^2}{\sum_{i=1}^{50} a_{ti}^2 + \sum_{i=1}^{50} p_{ti}^2}
\]

(16)

5. Empirical comparison and discussion

5.1. Comparison of predictive accuracy on PPA

In this section, the predictive abilities of the distance-based CTSP predictors are compared. The actual EEP values at 0.1 ms, 0.2 ms, \ldots, 50 ms time points of target case and the corresponding predictive EEP values generated by BTWD, ED, MD, DTW and TWED are depicted in Figs. 7 and 8, so there are total 500 \( \times \) 5 predictive values in one time hold-out validation. Then, in terms of the predictive results, we will be able to calculate PPA to examine the superiority of BTWD compare with other comparative approaches. Table 3 shows PPA values of each comparative method at 0.1 ms, 1 ms, 5 ms, 10 ms, 20 ms, 30 ms, 50 ms time points, along with their mean and standard deviation (SD) statistical indices. From Table 3, we can find BTWD achieves higher predictive accuracy than ED, MD, DTW, and TWED by 0.1086, 0.1217, 0.0642 and 0.043 respectively on mean index, and the PPA behavior of BTWD is relatively steadier than others with lower SD value by \(-0.0597, -0.0972, -0.054\) and \(-0.0794\).

5.2. Comparison of predictive accuracy on SPA

As shown in Figs. 7 and 8, the path of EEP elicited curve produced by BTWD is approximate to the actual EEP curve of target case. On the other hand, there are big differences between actual results and predictive results from ED and MD predictors in latter time segments. Conversely, the differences between actual results and predictive results from DTW and TWED are relatively big in early time segments. We can also come to the same conclusion from the observation of comparison on SPA index among five predictors. In this empirical comparison, we divided the time series [0.1 ms, 500 ms] into six segments, namely, [0.1 ms, 1 ms], [1 ms, 5 ms], [5 ms, 10 ms], [10 ms, 20 ms], [20 ms, 30 ms] and [30 ms, 50 ms], and SPA value can be figured out in each temporal segment. Table 4 lists the SPA values of comparative methods in six segments. In former 0.1 ms to 10 ms time series, BTWD outperforms DTW and TWED performances by (0.0503, 0.0455), (0.0847, 0.061) and (0.091, 0.0864), but marginally outperforms ED and MD performances by \((-0.0106, 0.0096), (0.0013, 0.0006)\) and \((0.0026, 0.0349)\) respectively in [0.1 ms, 1 ms], [1 ms, 5 ms] and [5 ms, 10 ms] segments. While in later 10–50 ms time series, it seems to be the opposite, because BTWD marginally outperforms DTW and TWED performances by \((0.0111, -0.0191), (0.0501, 0.0082)\) and \((0.0167, 0.0074)\), and outperforms ED and MD performances by \((0.1145, 0.1436), (0.1947, 0.1935)\) and \((0.1762, 0.1671)\) in [10 ms, 20 ms], [20 ms, 30 ms] and [30 ms, 50 ms] segments. This phenomenon provides some evidences on the assumption that the effectiveness of various distance metrics for predictive results in different temporal periods is different. Overall, the BTWD metric yields to outperform other distance metrics in the process of EEP prediction throughout [0.1 ms, 50 ms] period, which outperforms other four predictive methods by 0.0798, 0.0916, 0.0507 and 0.0316 on SPA mean index, as shown in Table 4.

5.3. Comparison of predictive accuracy on TIC

The TIC movements produced by BTWD, ED, MD, DTW and TWED are displayed in Fig. 9, from which BTWD achieves the best performance among other predictors, because the mean index of TIC on 30 hold-out data created by BTWD, ED, MD, DTW and TWED are 0.1645, 0.2096, 0.2101, 0.1804 and 0.1898, respectively. It indicates that BTWD outperforms other four comparative predictors on TIC by 0.0451, 0.0456, 0.0159 and 0.0253. Furthermore, only one TIC value of BTWD is higher than 0.3, as TIC value below 0.3 means a good agreement with actual data \([73,74]\), while some TIC values of ED and MD, DTW and TWED are over 0.3. It confirms that the EEP prediction using BTWD predictor adequately predicts the activity of the eliciting EEPs under these experimental conditions. On the other hand, it needs to notice that the SDs of TIC for five predictors are 0.0745, 0.0802, 0.0961, 0.0660 and 0.0618, so BTWD is outperformed by time series distance metrics, which demonstrates that we should adopt more advanced optimization algorithms for \(p, \omega, \lambda, \mu\) and \(k\) to improve the stability of BTWD performance.

5.4. Comparison of computational complexity between BTWD and time series distances

The computational complexity is another important criterion for evaluating time series distance method \([59,61]\), especially when the dataset is large with more features and more cases \([19]\). It has been established in \([50,52,53,58–60]\) that the complexities of DTW, TWED, ERP and SEA are \(O(n \times m), O(n \times m)\),
$O(n \cdot m + n \log_2(n) + m \log_2(m))$ and $O(2(n^2 + m^2) + 2(n + m))$ respectively, where $n$ and $m$ are the lengths of two time series being matched. In basis of the mechanisms of ED and TWED, the computational complexity of our proposed BTWD is $O(n \cdot m + n + m)$ for $\Theta(\tilde{a}_j \rightarrow \tilde{b}_j)$ match operation, and the costs of $\Theta(\tilde{a}_j \rightarrow A)$ and $\Theta(\tilde{b}_j \rightarrow A)$ operations can be tabulated to speed up the calculation leading to $O(2n + 2m)$ complexity. For EEP prediction, the length of each EEP time series sequence is equal, i.e., $n = m$, then the complexities of BTWD, DTW, TWED, ERP and SEA are $O(n^2 + 2n)$, $O(4n)$, $O(n^2)$, $O(n^2 + 2n \log_2(n))$, and $O(4n^2 + 4n)$. The relevant mathematic
Fig. 10. Computational complexities comparison between BTWD and other time series distances.

graphs for \( n = 10, 11, \ldots, 50 \) are depicted in Fig. 10, from which it shows that BTWD is marginally outperformed by DTW and TWED, and significantly outperforms ERP and SEA for from the aspect of computational complexity.

6. Conclusion

This study constructed a case based time series prediction method (CTSP) by using new distance metric, which demonstrated good predictive performance in EEG prediction of optic nerve visual prostheses research. As EEP experimental case contains the stationary electrical stimulation values and time-varying EEG elicited values, we proposed a biased time wrapping distance (BTWD) metric in similarity measure of case retrieval process, which was composed of stimulation set difference (Diff\( J \)) and EEG sequence difference (Diff\( J \)). Meanwhile, the time-independent bias configuration function was added into BTWD function to illustrate the different influences of Diff\( J \) and Diff\( J \) at various time points for predictive results. After that, the adaptation coefficient (AC) of each similar case was calculated, and similarity-related adaptation coefficient summation was employed to yield the predictive EEG values in principle of \( k \) nearest neighbors. Finally, we gave predictive performance comparison to investigate the superiority of EEP predictor using BTWD metric. From the results of empirical comparison, we can find that our proposed method produces superior performance in EEG prediction from the perspective of predictive accuracy and computational complexity.

This research also has some limitations. Conclusions drawn out are on the foundation of the experiment. Further discussions should be considered under other validation techniques, e.g., the resubstitution method, McNemar test and t-statistics test. In addition, more EEP experimental data should be collected, and the global parameter optimization and data weighting methods should be researched to improve the efficiency of BTWD metric. Finally, besides in biomedical field, we believe that the proposed method is suitable for TSP in other practical applications areas, e.g., stock price or other real-world time-series problems, where BTWD metric could be obtained by figuring out the input parameter set difference (Diff\( J \)) and output time sequence difference (Diff\( J \)). So the applicability of BTWD to data from other time series domains should be investigated in future.

Acknowledgements

This research is supported by the National Basic Research Program of China (2011CB707503 and 2011CB013305). China Postdoctoral Science Foundation funded project (2012M510112). The authors are grateful to the editors and the anonymous reviewers for their insightful comments and suggestions.

References


