Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)

An Efficient Time Series Subsequence Pattern Mining and Prediction Framework with an Application to Respiratory Motion Prediction

Shouyi Wang, Kin Ming Kam
Department of Industrial, Manufacturing & Systems Engineering, University of Texas at Arlington, TX 76019

Cao Xiao,* Stephen Bowen,+ W. Chaovalitwongse*
*Department of Industrial and Systems Engineering
+Department of Radiation Oncology,University of Washington, Seattle, WA 98195

Abstract

Traditional time series analysis methods are limited on some complex real-world time series data. Respiratory motion prediction is one of such challenging problems. The memory-based nearest neighbor approaches have shown potentials in predicting complex nonlinear time series compared to many traditional parametric prediction models. However, the massive time series subsequences representation, the similarity distance measures, the number of nearest neighbors, and the ensemble functions create challenges as well as limit the performance of nearest neighbor approaches in complex time series prediction. To address these problems, we propose a flexible time series pattern representation and selection framework, called the orthogonal-polynomial-based variant-nearest-neighbor (OPVNN) approach. For the respiratory motion prediction problem, the proposed approach achieved the highest and most robust prediction performance compared to the state-of-the-art time series prediction methods. With a solid mathematical and theoretical foundation in orthogonal polynomials, the proposed time series representation, subsequence pattern mining and prediction framework has a great potential to benefit those industry and medical applications that need to handle highly nonlinear and complex time series data streams, such as quasi-periodic ones.

Introduction

Although various time series modeling methods have been proposed, it is still a challenge to make accurate and robust predictions for highly nonlinear and complex time series, such as respiratory motion trajectories. A few studies showed that the memory-based nearest neighbor approach can be useful to predict highly nonlinear and complex time series patterns (Keogh, Lin, and Fu 2005; Sasu 2012; Yankov, Decoste, and Keogh 2006). In general, a K-nearest neighbor (KNN) method can be used to identify K most similar time series patterns first and then predict the future pattern of a query time series. Though promising for predicting complex nonstationary time series, a significant limit of these methods is on efficient subsequence pattern searching in a low-dimensional space. Also the performance of these methods are sensitive to the number of nearest neighbors K, the ensemble functions that combine nearest neighbor, and the sliding window length. Currently, no effective approaches exist to guide the selection of these model parameters. If too many neighbors are selected, the predicted values of a query time series tend to be biased by including intrinsically different patterns; while if too few neighbors are used, the predictions tend to be influenced by noises and outliers with large variance. Such bias-variance dilemma is a well-known issue in KNN applications (T. Hastie and Friedman 2009), and no systematic approach has been developed to improve the efficiency of time series subsequence pattern mining and the robustness of nearest-neighbor approaches in time series prediction. In addition to pattern searching approaches, an efficient representation of time series itself is also critical in time series pattern recognition and prediction. There has been a huge interest in approximating time series with reduced dimensionality. Methods include discrete Fourier transformation (DFT), single value decomposition (SVD), discrete wavelet transformation (DWT), piecewise approximation, and etc. (Fu 2011). Take the popular DFT-based representation for example, although it can achieve accurate time series approximation using a few signature DFT coefficients, the representation cannot be applied to weighted/penalized pattern matching problems. In summary, four fundamental problems exist in time series pattern mining and are in urgent need to be solved, including 1) data representation: how to represent the essential shape characteristics of a time series pattern in a largely reduced dimensional space, while allows convenient penalization and subsequence pattern mining; 2) similarity measure and pattern matching: how to search perceptually similar patterns from massive time series data effectively and efficiently ; 3) parameter selection: how to develop a data-driven method to select parameters and avoid the bias-variance dilemma; 4) ensemble function: how can we ensemble in order to achieve reliable and robust predictions.

This study is motivated to address the aforementioned time series data mining challenges, and provide a solution to make accurate and robust predictions for complex time series data. In particular, we developed a novel time series pattern representation using orthogonal polynomial approximation. A set of theories and similarity distance measures were developed to achieve fast and efficient time series pattern mining with flexible temporal weighting capability in massive time series subsequences. Based on the new time series representation, we further developed an orthogonal-polynomial-based variant-nearest-neighbor (OPVNN) approach to tackle the challenges of highly complex time series prediction. The proposed OPVNN method has been successfully applied to the challenging respiratory motion prediction problem, as most of the current time series models and machine learning algorithms still cannot achieve satisfactory prediction performance for clinical applications. The experimental results show that the proposed method achieved significantly superior and more robust prediction performance than the state-of-the-art prediction methods on respiratory motion trajectories from 27 patients.

The rest of the paper is organized as follows. In Section **??**rep we introduce the mathematical formulations of orthogonal polynomial based time series representation, and the related similarity measures using orthogonal polynomial coefficients. In Section **??**framework we present the proposed OPVNN prediction framework for multi-step time series prediction. Then in Section **??**result we provide the experimental results of the proposed method for the challenging respiratory motion prediction problem. Finally, we conclude the paper in Section **??**conclude.

Orthogonal-Polynomial Based Time Series Representation and Pattern Mining

Given a real-valued time series $\boldsymbol{x} = [x_0, x_1, \ldots, x_N]$, we assume $\boldsymbol{x}$ can be modeled by a parameterized function $f(t) : \mathbb{R} \rightarrow \mathbb{R}$. In addition, we assume that the function $f(t)$ can be represented by a linear combination of $K+1$ basis functions f_k with a parameter vector $\boldsymbol{w}$ with elements $w_k \in \mathbb{R}$ in the following form:

$$f(t) = \sum_{k=0}^{K} w_k \cdot f_k(t), \tag{1}$$

where the basis function f_k can be polynomials, sigmoid functions, wavelets, etc. The values of the $K+1$ basis functions for the $N+1$ time series points can be written into a matrix $\boldsymbol{F} \in \mathbb{R}^{(N+1)(K+1)}$ as follows:

$$\boldsymbol{F} = \begin{pmatrix} f_0(t_0) & \cdots & f_K(t_0) \\ \vdots & \ddots & \vdots \\ f_0(t_N) & \cdots & f_K(t_N). \end{pmatrix} \tag{2}$$

Then the linear least-squares approximation problem for the time series $\boldsymbol{x}$ using the function f can be denoted by

$$\min_{w} \|\boldsymbol{F}\boldsymbol{w} - \boldsymbol{x}\|^2, \tag{3}$$

where $\|\cdot\|$ is the Euclidean norm. The least-squares solution $\boldsymbol{w}_{LS}$ can be obtained by setting the derivative of the above objective function with respect to the parameter vector $\boldsymbol{w}$ to zero. And

$$\boldsymbol{w}_{LS} = (\boldsymbol{F}^T\boldsymbol{F})^{-1}\boldsymbol{F}^T\boldsymbol{x} = \boldsymbol{F}^{+}\boldsymbol{x}, \tag{4}$$

where $\boldsymbol{F}^{+} = (\boldsymbol{F}^T\boldsymbol{F})^{-1}\boldsymbol{F}^T$ is the pseudo-inverse $\boldsymbol{F}$. If we select orthogonal polynomials as the basis functions, such that $\sum_{n=0}^{N} f_{k_1}(t_n) \cdot f_{k_2}(t_n) = 0$ for any two basis functions f_{k_1} and f_{k_2} with $k_1 \neq k_2$, then we have the following property of $\boldsymbol{F}$

$$\begin{aligned} \boldsymbol{F}^T\boldsymbol{F} &= \begin{pmatrix} \|f_0\|^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \|f_K\|^2 \end{pmatrix} \\ &= \Sigma_F \end{aligned} \tag{5}$$

With orthogonal basis functions, the least-squares solution of the parameter vector $\boldsymbol{w}$ can be rewritten as

$$\begin{aligned} \boldsymbol{w}_{LS} &= (\boldsymbol{F}^T\boldsymbol{F})^{-1}\boldsymbol{F}^T\boldsymbol{x} \\ &= \begin{pmatrix} \sum_{n=0}^{N} \frac{x_n}{\|f_0\|^2} f_0(t_n) \\ \vdots \\ \sum_{n=0}^{N} \frac{x_n}{\|f_K\|^2} f_K(t_n) \end{pmatrix} \end{aligned} \tag{6}$$

Then the sum of squared error SSE_{LS} of the orthogonal polynomial approximation by

$$SSE_{LS} = \sum_{n=0}^{N} x_n^2 - \sum_{k=0}^{K} w_k^2 \|f_k\|^2. \tag{7}$$

As a time series may drift over time and the relative morphological patterns are more informative in similarity pattern searching, one can remove the "drift" of a time series by removing its mean before approximation. Using the concept of coefficient of determination, often denoted by R-squared, we can introduce a new approximation accuracy measure of the time series $\boldsymbol{x}$ by

$$\begin{aligned} R_{OP}^2 &= 1 - \frac{SSE_{LS}}{SST(\boldsymbol{x})}, \\ &= \frac{\boldsymbol{w}_{LS}^T \Sigma_F \boldsymbol{w}_{LS}}{\boldsymbol{x}^T\boldsymbol{x}}. \end{aligned} \tag{8}$$

Where the $SST(\boldsymbol{x}) = (\boldsymbol{x} - \bar{\boldsymbol{x}})^T(\boldsymbol{x} - \bar{\boldsymbol{x}}) = \boldsymbol{x}^T\boldsymbol{x}$ for the zero-mean time series vector, and Σ_F is the pre-computed $(K+1) \times (K+1)$ diagonal matrix, and K is the maximum order of orthogonal polynomials used for approximation. With R_{OP}^2 in formula 8, one can obtain the approximation performance of any time series efficiently only using the orthogonal polynomial coefficients and the inner product of the time series vector $\boldsymbol{x}$.

The Benefits of Orthogonal Polynomials Approximations:

- **Fast updating and approximation**: according to equation 6, the least-squares solution of the parameter $\boldsymbol{w}_{LS}$ is simply a weighted linear combination of the time series data values. This property enables a super fast approximation for real-time time series streams with a very low computational complexity.
- **Efficient control of approximation performance**: due to the orthogonality, the coefficients of different polynomial orders are independent to each other. The coefficients of lower order terms do not change as the order of approximation polynomials goes up. If lower order approximation is needed (for example, avoid over-fitting), we can

just set the coefficients of higher order terms to zero without re-fitting the approximation model. The approximation accuracy can be conveniently controlled by the proposed measure in equation 8.

- **Sparse representation**: the coefficient vector of orthogonal polynomial approximation is a high-level sparse representation of raw time series data. The coefficient vector retains the essential time series pattern information in a low-dimensional space. With attractive mathematical properties, one can achieve very efficient similarity pattern mining in the coefficient space of the time series data.

Similarity Distance Measure

In time series similarity pattern matching, different temporal points may have different levels of importance in pattern searching. For example, we may consider the most recent time points are more important than the older ones to search for similar time series subsequence patterns. To enforce different similarity-matching requirement for different time points, we introduce a error-penalty vector in solving the least-squares problem defined in 3.

***Theorem* 1.** *Let $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ be two time series with a length of $N+1$ time points, $\hat{\boldsymbol{x}}_1$ and $\hat{\boldsymbol{x}}_2$ are the Kth order orthogonal polynomial approximations of the two time series with coefficient vectors $\boldsymbol{w_1}$ and $\boldsymbol{w_2}$, respectively. Given an error-penalty vector $\boldsymbol{P} = p_0, p_1, \ldots, p_N$ for the $N+1$ time series points. The weighted Euclidean distance between the two time series approximations $\hat{\boldsymbol{x}}_1$ and $\hat{\boldsymbol{x}}_2$ can be obtained from the orthogonal polynomial coefficients directly in $\boldsymbol{O(K^2)}$.*

Proof. We can rewritten the penalty vector $\boldsymbol{P}$ in a diagonal matrix form $\boldsymbol{P_d}$. Then the weighted sum of squared residual errors $WSSE$ between $\hat{\boldsymbol{x}}_1$ and $\hat{\boldsymbol{x}}_2$ can be represented by:

$$\begin{aligned} WSSE(\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2) &= \boldsymbol{(w_1 - w_2)^T F^T P_d F (w_1 - w_2)} \\ &= \boldsymbol{(w_1 - w_2)^T G_F (w_1 - w_2)} \\ &= \boldsymbol{[w_1^T G_F w_1, -2 w_1^T G_F][1, w_2^T]^T} \\ &\quad \boldsymbol{+ w_2^T G_F w_2} \\ &= \boldsymbol{H(w_1, G_F)[1, w_2^T]^T + w_2^T G_F w_2}. \end{aligned} \tag{9}$$

The given error-penalty vector P generate a $K \times K$ square matrix $\boldsymbol{G_F}$. Using formula 9, the weighted Euclidean distance can be obtained by the orthogonal polynomial coefficients directly in $\boldsymbol{O}(K)$. □

***Theorem* 2.** *Now assume we have a pattern library $\boldsymbol{B}$ with orthogonal polynomial coefficient vectors of M time series subsequences. For a query time series $\boldsymbol{x}_t$, the similarity pattern searching problem in the pattern library can be efficiently formulated by the inner product of the transformed matrix $\boldsymbol{H(B, \Sigma_F)}$ or $\boldsymbol{H(B, G_F)}$ and the query orthogonal polynomial coefficient vector $\boldsymbol{w}_t$.*

Proof. For the equal-penalty case, the Euclidean distances between the approximation of the query time series and the library patterns can be represented by

$$SSE(\boldsymbol{B}, \boldsymbol{x_t}) = \boldsymbol{H(B, \Sigma_F)[1, w_t^T]^T + w_t^T \Sigma_F w_t}. \tag{10}$$

Here we notice that the second term $\boldsymbol{w}_t^T \Sigma_F \boldsymbol{w}_t$ is only related to the query time series itself, this part can be removed in the similarity pattern search in the library. Thus, we can define a new distance measure by

$$D(\boldsymbol{B}, \boldsymbol{x_t}) = \boldsymbol{H(B, \Sigma_F)[1, w_t^T]^T}. \tag{11}$$

With this formulation, the distances between the approximated query time series and the library patterns can be computed efficiently by the inner product of the transformed matrix $\boldsymbol{H(B, \Sigma_F)}$ and the query orthogonal polynomial coefficient vector $\boldsymbol{w}_t$. The computational complexity is in $\boldsymbol{O}(MK)$, and is not affected by the time series data length N. Similarly for the weighted-penalty case, the new distance measure can be obtained by

$$D_w(\boldsymbol{B}, \boldsymbol{x_t}) = \boldsymbol{H(B, G_F)[1, w_2^T]^T}. \tag{12}$$

The weighted Euclidean distance between the query time series and the library patterns can also be obtained efficiently in $O(MK)$ without any extra computational load. □

Based on the new similarity distance formulation in 11 and 12, we can construct pattern library using the transformed coefficient vector using the transformation function $\boldsymbol{H}$. The distance vector for a query time series $\boldsymbol{x_t}$ can be obtained efficiently by the inner product of the enhanced orthogonal polynomial coefficient vector $[1, \boldsymbol{w}_t]$ and the pattern library matrix. The transformed vector $\boldsymbol{H(w_t, \Sigma_F)}$ or $\boldsymbol{H(w_t, G_F)}$ is then stored in the pattern library instead of original coefficients.

Proof. Assume the sliding window to construct the pattern library has a length of $N+1$, a query time series $\boldsymbol{x'} = [x_0, x_1, \ldots, x_{L-1}]$ has a length of L with $L < N+1$, and the corresponding orthogonal polynomial coefficient vector is $\boldsymbol{w}'$. To search for similar patterns in the pattern library, we can design a error-penalty weight vector $P' = [p_0, p_1, \ldots, p_N]$ such that L consecutive penalty weights are non-zero and the remaining penalty weights are zeros. As the library stores time series subsequences continuously with a small moving step, we can just set the the first L penalty weights non-zero and others zero. We denote the diagonal matrix form of the penalty vector by P'_d. Then we can employ the formula 9 to calculate the distances between the query time series and the subsequences of length L in the library patterns. In particular, denote $\boldsymbol{F^T P'_d F = G'_F}$, the resulting distance vector can be obtained by

$$D_w(\boldsymbol{B}, \boldsymbol{x'}) = \boldsymbol{H(B, G'_F)[1, w'^T]^T}. \tag{13}$$

□

Statistical Similarity Measure in Pattern Matching: To conveniently quantify the similarity between two time series patterns, the sum of squared residual errors (SSE) is commonly used. However, the value of SSE is dependent on time series amplitude and data length. To automatically scale the concept of similarly for massive time series with different values ranges, we propose a statistical measure as the similarity criterion to determine the best-matching similar patterns automatically without giving a fixed number of neighbors in pattern searching. Assume a pattern library with n subsequences $\hat{\boldsymbol{x}}_j$ with $j = 1, 2, \ldots, n$, and the approximation of a query time series is $\hat{\boldsymbol{x}}_t$. The weighted similarity measure of two time series sequences is defined by

$$\begin{aligned} R^2_{wm}(\hat{\boldsymbol{x}}_j, \hat{\boldsymbol{x}}_t) &= 1 - \frac{WSSE(\hat{\boldsymbol{x}}_j, \hat{\boldsymbol{x}}_t)}{WSST(\hat{\boldsymbol{x}}_t)} \\ &= 1 - \frac{\boldsymbol{H(w_j, G_F)[1, w_t^T]^T + w_t^T G_F w_t}}{\boldsymbol{w_t^T G_F w_t}} \\ &= -\frac{\boldsymbol{H(w_j, G_F)[1, w_t^T]^T}}{\boldsymbol{w_t^T G_F w_t}}. \end{aligned} \tag{14}$$

With the above definition, $R^2_m(\hat{\boldsymbol{x}}_j, \hat{\boldsymbol{x}}_t)$ or $R^2_{wm}(\hat{\boldsymbol{x}}_j, \hat{\boldsymbol{x}}_t)$ is closer to one if two time series patterns are more similar to each other. If the two time series patterns are identical, $R^2_m(\hat{\boldsymbol{x}}_j, \hat{\boldsymbol{x}}_t)$ or $R^2_{wm}(\hat{\boldsymbol{x}}_j, \hat{\boldsymbol{x}}_t)$ is equal to 1. With such a nice property, one can set the similarity requirement conveniently, such as 0.95, in the similarity pattern searching process. The patterns with higher similarity values are selected as best-matching patterns in the pattern library.

Online Time Series Subsequence Mining & Prediction Framework

The memory-based nearest neighbor approaches have potentials to predict highly complex and nonlinear time series than many traditional parametric time series models. However, it is still challenging to mining massive time series subsequences efficiently. The selection of time series representations in low-dimensional space, the selection of similarity distance measures, the determination of the number of nearest neighbors, and the ensemble functions of nearest neighbors are still serious issues that limit the performance of the current nearest neighbor approaches for prediction of highly nonlinear and complex time series patterns. In this study, we propose a new approach to address these problems, called orthogonal polynomial based variant nearest neighbor (OPVNN) prediction framework as shown in the following subsections.

Time Series Prediction Problem Statement

For a real-time time series $\boldsymbol{y} = [y_1, y_2, \ldots, y_t, \ldots]$, a sliding window is applied to monitor the time series with a window length of L_{win} time points and a step size of L_{step} time points. Given a prediction horizon of h steps, at each time step t, the prediction task is to find variant k best-matching reference patterns in the time series history, and ensemble the h-step-ahead values of the reference patterns $[y_{r1,h}, y_{r2,h}, \ldots, y_{rk,h}]$ to make the h-step prediction of the query time series at time point t using an ensemble function Θ. The prediction formulation can be written as

$$\hat{y}_{t+h} = \Theta(y_{r1,h}, y_{r2,h}, \ldots, y_{rk,h}). \tag{15}$$

Time Series Subsequence Representation

As the sliding window monitors the time series $\boldsymbol{x}$, each time series subsequence with a length of L_{win}. At time point t, denote the window time series subsequence by $\boldsymbol{y}_t = [y_{t-L_{win}+1}, y_{t-L_{win}+2}, \ldots, y_{t-1}, y_t]$. The orthogonal polynomial approximation is applied to the zero-mean vector $\tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t - \bar{\boldsymbol{y}}_t$, the resulting approximation coefficient vector is $\boldsymbol{w}_t$. According to distance formulas 9, we transform the original coefficient vector into a new space by the transformation function $\boldsymbol{H}(\boldsymbol{w}_{rt}, \Sigma_F)$ for equal-penalty matching case, or $\boldsymbol{H}(\boldsymbol{w}_{rt}, G_F)$ for the weighted-penalty matching case. The pattern library stores the transformed coefficient vector of each time series subsequence.

Pattern Baseline Calibration

In addition to the zero-mean pre-processing, one can also apply different aligning methods for time series subsequences depend on applications. For example, for the respiratory time series prediction study, we noticed that the right-aligned pattern matching can be more effective to find best reference patterns for an accurate motion prediction. Instead of subtracting the mean, each subsequence is subtracted by its right-end value such that $\tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t - \boldsymbol{y}_t(L_{win})$. Then the orthogonal polynomial approximation is performed for the right-aligned pattern and store the transformed coefficient vector into the pattern library.

Adaptive Nearest Neighbor Searching

Instead of selecting a pre-determined number of nearest neighbors, we propose to employ the statistical similarity measures R_{wm} to identify best-matching neighbors for query time series patterns. The similarity measure is a scaled metrics defined in $[-\infty, 1]$. The closer the R_m or R_{wm} to one, the two comparing time series patterns are more similar. A lower value of R_m and R_{wm} indicates that the two time series patterns are more dissimilar. Based on such interpretable and easily controllable similarity measure, we can define the pattern-matching similarity by giving a threshold for similarity requirement. For example, we can set the similarity requirement at 0.95, such that $R^*_m = 0.95$. Then all the time series subsequences that meet the similarity-matching requirement can be considered as nearest neighbors of a query time series. In this way, we achieve the desirable data-driven variant-nearest-neighbor selection, and avoid the drawbacks of using a fixed number of nearest neighbors. It is noted that the similarity measure 14 is defined based on the zero-mean time series subsequences. If we use the right-aligned subsequence representation, one can just replace the denominator of the two similarity formulas by $SST(\hat{\boldsymbol{x}}_t) = (\boldsymbol{F}\boldsymbol{w}_t - \bar{\boldsymbol{x}}_t)^T(\boldsymbol{F}\boldsymbol{w}_t - \bar{\boldsymbol{x}}_t)$, all other parts are the same.

Statistical Screening of Predictive Neighbors

Assume we have a set of k best-matching reference subsequences, denoted by $\boldsymbol{R_I}$, that satisfy the similarity requirement R^*_m or $R_m w^*$. The next step is to make a statistical analysis of the h-step-forward values of the reference subsequences, denoted by $\boldsymbol{y}_r(t+h) = [y_{r1,t+h}, y_{r2,t+h}, \ldots,$

$y_{rk,t+h}$]. We estimate the distribution of the h-step-forward values, and denote the qth percentile of the distribution of the h-step future values $\boldsymbol{y}_r(t+h)$ by $P_{\boldsymbol{y}_r,q}$. Then we can determine the final predictive reference subsequence set $\boldsymbol{R_{II}}$ by removing the outliers in $\boldsymbol{y}_r(t+h)$ using the following statistical outlier detection formula:

$$\boldsymbol{R_{II}} = \{j \in RP_I | y_{rj,t+h} \leq P_{\boldsymbol{y}_r,75} + 1.5(P_{\boldsymbol{y}_r,75} - P_{\boldsymbol{y}_r,25}), \\ y_{rj,t+h} \geq P_{\boldsymbol{y}_r,75} + 1.5(P_{\boldsymbol{y}_r,75} - P_{\boldsymbol{y}_r,25})\}. \tag{16}$$

Similarity-Weighted Ensemble Function for Multi-Step Time Series Prediction

At time point t, assume we obtain s reference subsequence patterns after outlier removal. The h-step future time series value y_{t+h} can be predicted by combining the h-step future values of the reference patterns using an ensemble function $\Theta(y_{r1,h}, y_{r2,h}, \ldots, y_{rs,h})$. As there is no existing guidelines for ensemble functions, a most common way of current approaches is to take the mean of the h-step future values of the reference patterns (Ichiji et al. 2013a; 2013b). In this study, we propose a similarity-weighted ensemble function to combine the future values of the reference subsequences $y_{r1,h}, y_{r2,h}, \ldots, y_{rs,h}$. In particular, the similarity-weighted ensemble function for the weighted-penalty case can be defined by replacing R_m^2 with the weighted R-squared measure R_{wm}^2 in the following

$$\hat{y}(t+h) = \frac{\sum_{j=1}^{s} R_{wm}^2(\boldsymbol{y}_{rj})(\boldsymbol{y}_{rj,h} - o_j)}{\sum_{j=1}^{s} R_{wm}^2(\boldsymbol{y}_{rj})} + o_t. \tag{17}$$

Experimental Results

The proposed OPVNN method was evaluated by the challenging respiratory motion prediction problem and compared with the state-of-the-art time series prediction methods.

Data Acquisition

Time series of abdominal displacement of 27 lung and liver cancer patients were collected with the Real-time Position Management$^{\text{TM}}$(RPM)(Varian Inc., Santa Clara, CA) infrared camera and reflective marker block system during their PET/CT examination. These time series trajectories serve as respiratory motion surrogate of lung tumor inside of the body (Wang et al. 2014). The sampling rate of the respiratory traces was 30 Hz. The duration of data collection for each patient was from 30 to 45 minutes when they did PET/CT scan.

Baseline Methods

This study was motivated from our research project of respiratory motion management in radiotherapy. The current prediction methods still cannot work well for real-time prediction of respiratory motion in clinics. To tackle this problem, we studied various different methods, including Gaussian Process methods, Kalman Filter, Kernel density estimation (KDE), support vector regression (SVR), Seasonal ARIMA, Neural Networks, etc. Respiratory motion prediction is a challenging unsolved problem, Ernst et al. (Ernst et al. 2013) made comprehensive comparisons on various existing prediction methods, including the above mentioned ones, the wLMS and SVR were the top performers, we had similar conclusions in our experiments. Also we identified a most recent method res-TVSAR from Ichiji et al. (Ichiji et al. 2013a; 2013b) that performed better than KDE and Seasonal ARIMA. This paper compared the proposed method with the carefully selected most recent top performing models based on our extensive investigation on various methods. The prediction performances are presented in the following.

Experimental Design and Settings

We compared our method with the state-of-the-art time series pattern mining and prediction approaches, including wLMS, SVRpred and TVSAR, and Seasonal ARIMA. To make a fair comparison, the parameters of all the prediction approaches were determined through a training-testing process. In particular, we used the first 15 minutes of the respiratory trajectory of each patient as the training data for parameter optimization. In particular, 80% of the training data were used for construct a pattern library and the remaining 20% of the time series data was used for prediction performance testing. The optimal parameter setting is the set that maximizes the testing accuracy. The unused next 15 minutes of the respiratory data of each patient was used for prediction performance validation. We report the validation results of each method over the 27 patients.

For the OPVNN method, the personalized sliding window size was selected from the options of [0.50, 0.75, 1.0, 1.25, 1.50] times of the median length of respiratory cycles of each patient. The step length was three time points representing 100ms in respiratory motion. A weighted-penalty vector was employed in the pattern matching. Since we assign higher priority to the sub-pattern closer to the prediction point in similar-pattern searching, as shown in Figure 1, we put higher matching-error penalties to the right side (more recent time points). Setting the long-horizon matching-error penalty (left-side) to 1 as a baseline, we made the recent short-horizon penalty (right-side) as a parameter with options of [1, 5, 10]. The length of the high-penalty portion was selected from [0.1, 0.15, 0.2, 0.3] of the sliding window. These parameter were all determined in the training-testing process. In the adaptive nearest neighbor searching step, the similarity threshold R_{wm}^* was set at 0.95. For a sliding window query time series, if none of the library patterns meet the similarity requirement 0.95, we reduce the similarity requirement by 0.02, until at least three reference subsequences are selected for prediction. The similarity-weighted ensemble function defined in equation 17 was applied to make multi-step predictions for each sliding window.

Prediction Performance Criterion

The prediction performance was evaluated by a R-squared measure denoted by

$$R_{pred}^2 = 1 - \frac{SSE(\boldsymbol{y}_{val} - \boldsymbol{y}_{pred})}{SST(\boldsymbol{y}_{val})} = 1 - \frac{\|\boldsymbol{y}_{val} - \boldsymbol{y}_{pred}\|^2}{\|\boldsymbol{y}_{val} - \bar{\boldsymbol{y}}_{val}\|^2},$$

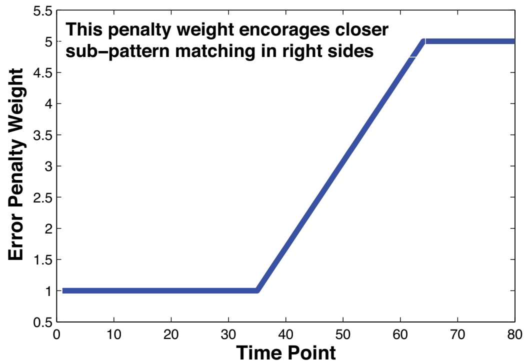

Figure 1: The employed matching-error penalty weight function used for similarity pattern searching.

where $\boldsymbol{y}_{val}$ is the actually time series data in validation, $\boldsymbol{y}_{pred}$ is the predicted values for a prediction horizon h. This measure evaluates how close the predicted respiratory motion $\boldsymbol{y}_{pred}$ to the actual motion trajectory $\boldsymbol{y}_{val}$ of each patient.

Prediction Performance Comparison

The proposed OPVNN framework consists of four key components, including the OP-based approximation, the adaptive-aligned similarity pattern matching, the time-delay-penalized variant-best-neighbor pattern searching, and the R^2-weighted ensemble prediction scheme. We evaluated the effectiveness of each key component by replacing the component with a widely used method while keeping the other three components unchanged. In particular, the four replacement frameworks include 1) replaced the OP-based time series approximation by the popular DFT representation developed in (Cooley and Tukey 1965); 2) employed the simple averaged prediction from best-neighbors instead of the R^2-weighted ensemble method; 3) applied equal weights to the entire time series pattern without penalties over time; 4) performed similarity pattern searching using raw time series patterns without adaptive alignment at each prediction step. Figure 2 shows the box plots of the five models with respect to prediction horizon of 10 steps. One can see clearly that the best prediction performance was achieved by the proposed OPVNN method with all the four key components were actively working together. Removing any of the four key components can reduce the overall prediction performance significantly.

Figure 3 shows the R^2_{pred} values of the proposed OPVNN method and the four state-of-the-art time series modeling and prediction methods for respiration motion prediction, including res_TVSAR (Ichiji et al. 2013a; 2013b), wLMS (Ernst, Schlaefer, and Schweikard 2007; Ernst et al. 2013), SVRpred (Ernst and Schweikard 2009; Ernst et al. 2013) and SARIMA (T. Hastie and Friedman 2009). It clearly showed that the proposed OPVNN approach achieved significant better prediction performance than the four state-of-the-art methods for all prediction horizons up to 30 steps. The prediction performance of the OPVNN method were also a lot more robust (with less prediction variances over the 27 patients) than the four comparing recent methods. This experimental result outcome further confirmed the high efficiency and effectiveness of the proposed OPVNN method for massive time series subsequence pattern mining and prediction of complex time series data (like respiratory motion trajectories).

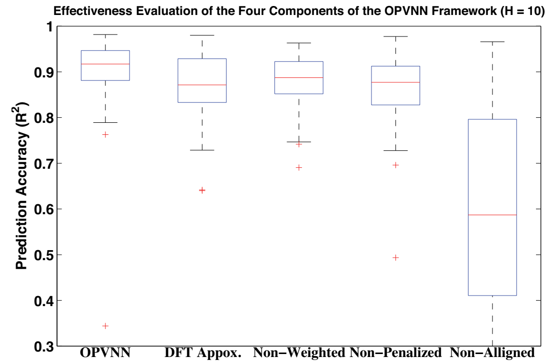

Figure 2: Effectivenessof the four key components of the OPVNN framework with a prediction horizon of 10 steps.

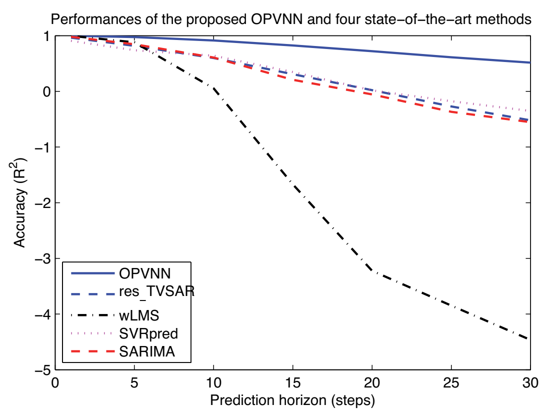

Figure 3: Prediction performance of the proposed OPVNN approach and the four state-of-the-art methods including res_TVSAR, wLMS, SVRpred and SARIMA.

Discussion and Conclusion

In this study, we developed a new time series modeling and representation method using orthogonal-polynomial approximations. A set of theories and similarity distance measures have been developed to achieve fast and efficient pattern mining for massive time series subsequences. Based on the OP-based representation, we proposed a new time series pattern mining method OPVNN to achieve an efficient and robust prediction of highly complex time series patterns. The proposed OPVNN method has showed its effectiveness in the challenging respiratory motion prediction problem, and achieved and significant superior and robust performance to most of the current time series prediction approaches. In summary, this study developed a new time series representation method using orthogonal polynomials, and a new prediction framework for time series subsequence modeling, similarity-pattern searching, and multi-step prediction. The proposed OPVNN approach is a general framework that has a great potential to benefit many industry and medical applications that need handling highly complex real-time time series data streams.

References

Cooley, J., and Tukey, J. 1965. An algorithm for the machine calculation of complex fourier series. *Math. Comput.* 19:297301.

Ernst, F., and Schweikard, A. 2009. Forecasting respiratory motion with accurate online support vector regression (svrpred). *International Journal of Computer Assisted Radiology and Surgery* 4(5):439–447.

Ernst, F.; Drichen, R.; Schlaefer, A.; and Schweikard, A. 2013. Evaluating and comparing algorithms for respiratory motion prediction. *Physics in Medicine and Biology* 58:3911–3929.

Ernst, F.; Schlaefer, A.; and Schweikard, A. 2007. Prediction of respiratory motion with wavelet-based multiscale autoregression. *Medical Image Computing and Computer-Assisted Intervention* 668–675.

Fu, T. C. 2011. A review on time series data mining. *Engineering Applications of Artificial Intelligence* 24(1):164 – 181.

Ichiji, K.; Homma, N.; Sakai, M.; Abe, M.; Sugita, N.; and Yoshizawa, M. 2013a. *A Respiratory Motion Prediction Based on Time-Variant Seasonal Autoregressive Model for Real-Time Image-Guided Radiotherapy*. INTECH. chapter Chapter 5, 75–90.

Ichiji, K.; Homma, N.; Sakai, M.; Narita, Y.; Takai, Y.; and Zhang, X. 2013b. A time-varying seasonal autoregressive model-based prediction of respiratory motion for tumor following radiotherapy. *Computational and Mathematical Methods in Medicine* 2013:Article ID 390325, 9 pages.

Keogh, E.; Lin, J.; and Fu, A. 2005. HOT SAX: efficiently finding the most unusual time series subsequence. *Proceedings of the 5th IEEE International Conference on Data Mining* 226–233.

Sasu, A. 2012. K-nearest neighbor algorithm for univariate time series. *Bulletin of the Transilvania University of Brasov* 5(2):147–152.

T. Hastie, R. T., and Friedman, J. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2nd edition. 37.

Wang, S.; Bowen, S. R.; Chaovalitwongse, W. A.; Sandison, G. A.; Grabowski, T. J.; and Kinahan, P. E. 2014. Respiratory trace feature analysis for the prediction of respiratory-gated pet quantification. *Physics in Medicine and Biology* 59(4):1027.

Yankov, D.; Decoste, D.; and Keogh, E. 2006. Ensembles of nearest neighbor forecasts. 545–556.