Wrapper Feature Extraction for Time Series Classification Using Singular Value Decomposition

Hui Zhang, Tu Bao Ho and Saori Kawasaki
School of Knowledge Science, Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan
{zhang-h, bao, skawasa}@jaist.ac.jp

ABSTRACT

Time series classification is an important aspect of time series mining. Recently, time series classification has attracted increasing interests in various domains. However, the high dimensionality property of time series makes time series classification a difficult problem. The so-called *curse of dimensionality* not only slows down the process of classification but also decreases the classification quality. Many dimensionality reduction techniques have been proposed to circumvent the *curse of dimensionality* problem for improving the time series classification performance. However, most of the proposed time series dimensionality reduction algorithms don't utilize the information of data labels that is crucial for the classification problem.

We propose a wrapper feature extraction algorithm incorporating with the classification algorithm for time series classification in this paper. The classification errors estimated by cross validation are taken as the measure of the quality of dimensionality reduction. As a set of univariate time series can be represented as a matrix, singular value decomposition is used as the feature extraction algorithm to approximate the original time series with a lower-rank matrix. By analyzing the characters of singular vectors for noisy data, we propose several efficient search algorithms. Comparison Experiments on several benchmark time series data validate the usefulness of the proposed approach.

Keywords: time series, feature extraction, wrapper approach, singular value decomposition

1. INTRODUCTION

Time series data made up of sequences of values changing with time. Time series data are popularly encountered in many domains such as finance, biomedicine, and bioinformatics. Recently, time series mining has attracted increasing interests and time series classification is one of the main directions of time series mining. Normally, time series data are high dimensional data, for example, if we collect stock market data everyday, the dataset containing one year data will has the dimensionality 365. The so-called *curse of dimensionality* problem makes time series classification a challenging problem. Curse of dimensionality refers to the situation that the number of instance must increase exponentially with the increase of dimensionality for a given level of classification accuracy [1]. For many real time series datasets, the number of instance is much less than that required for achieving high classification accuracy, thus dimensionality reduction algorithm is widely used as a preprocessing procedure for circumventing the curse of dimensionality problem. The additional benefits of dimensionality reduction are that it speeds the process of classification and improve the comprehensibility of the classification results [8].

Feature subset selection (FSS) and feature extraction (FE) are two commonly used techniques served for dimensionality reduction [3]. FSS selects a subset of features from original features directly. In comparison with FSS, FE creates new features based on transformation or combination of the original feature set. For the classification problem, FSS is based on a criterion to measure how strong the data labels associated with the features. The features with higher relationship with data labels are thought contribute more to classification algorithms. Many type of criteria have been proposed, such as mutual information [9] [13], nearest-neighbor label difference [10][11], correlation [12], etc. However, it has been recognized in FSS community recently that along with relative, the redundancy also affects the classification performance. The reason why high redundancy of features degrades the classification process by be explained with the mutual information theory mathematically [13]. The intuition behind this explanation is that if a set of features are associated together, then only one member of the set can represent the whole feature set. Several FSS methods have been introduced to reduce the redundancy among features and select the feature with minimized redundancy [13]-[15]. Nevertheless, we know there is strong relationship between every adhering data points within a time series, for example, today's stock market value has strong relationship with that of yesterday. This natural property is in conflict with the implying assumption that some features are

independent with that redundancy reduction can be performed. In this case, feature extraction techniques that allows to decorrelate the relation between adhering data points are much suitable for dimensionality reduction than applying feature subset selection directly to the time series data.

Many feature extraction algorithms have been proposed for time series classification, including Fourier transform [16], wavelets [17], Singular Value Decomposition [7], etc. However, to our knowledge, most of the proposed techniques are belonged to the unsupervised category without considering the information of data labels that is important for classification. In FSS community, using information of data labels for dimensionality reduction can be grouped into wrapper and filter approaches in terms of whether the classification algorithm is interwinded in the dimensionality reduction process [2]. Normally, wrapper approach that takes the performance of classification algorithm as a criterion for selecting the features can get higher classification accuracy than filter approach. In this paper, by borrowing the idea from feature selection to time series classification, we propose a wrapper feature extraction algorithm. Singular Value Decomposition (SVD) is an optimized orthogonal transform technique that allows for decomposing a time series data set into orthogonal subspaces. As the wrapper approach is slower than the filter approach, the crucial point of the wrapper approach is to design efficient search technique. We studied several heuristic search algorithms that can speed up the wrapper searching process.

The paper is organized as follows: Section 2 introduces the related work of using SVD in time series feature extraction. Section 3 presents our wrapper feature extraction algorithm. Experimental evaluation is given in Section 4. We conclude the paper with summarizing the main contributions in Section 5.

2. RELATED WORK

SVD is popularly used in time series querying. As *N*-nearest neighbor classification used in querying is also one of the basic technques of classification, the proposed algorithms can be viewed as belonged to time series classification. Korn et al. proposed using SVD for time series querying [7]. Ravi et al. proposed an efficient algorithm for saving the computational time of SVD with dynamic database by using the aggregate data from the existing data structure [4]. Chandrasekaran et al. introduced an updating SVD algorithm that is suitable for online learning to speed up the SVD process [5]. Castelli et al. clustered homegeneous data into groups firstly then separately reduced the dimensionality of each group using SVD [6].

To our knowledge, all the proposed methods fall into unsupervised feature extraction category don't use the information of data lables. Our approach that uses the classification accuracy as the measure for choosing the lower-rank approximation of the time series is different with all other proposed algorithms.

3. SVD BASED WRAPPER FEATURE EXTRACTION

3. 1. SVD Based Time Series Feature Extraction

For a time series dataset consisting *n* time series, where each time series has length *d*, we can represent the set of time series by a $n \times d$ matrix *A*. The matrix *A* with rank *k* can be decomposed using SVD by

$$A = U\Sigma V^T \tag{1}$$

A is the $n \times d$ data matrix composed of the *n* d-dimensional vectors, *U* is a $n \times n$ orthogonal matrix, and *V* is a $d \times d$ orthogonal matrix. *U* and *V* are called the left and right sigular vectors of *A*. Σ is a $n \times d$ diagonal matrix containing *k* singular values

$$\begin{bmatrix} s_1 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & s_k \end{bmatrix} \tag{2}$$

where $s_1 > s_2 > \ldots > s_k \geq 0$.

The SVD decomposition can also be written as

$$A = U\Sigma V^T = \sum_{i=1}^{k} s_i u_i v_i^T \tag{3}$$

where u_i is the *i*th column of *U* and v_i is the *i*th column of V. When using SVD for dimensionality reduction, the matrix *A* is approximated by a lower rank matrix *B* with which the error between *A* and *B* is minimized. It can be proven that the squred error minimizition refers to the situation that

$$B = \sum_{i=1}^{r} s_i u_i v_i \tag{4}$$

where r is the favorable rank, and s_i is the few largest singular values of *A* [18]. And the error between *A* and *B* is

$$\|A - B\| = \sum_{i=r+1}^{k} s_i u_i v_i \tag{5}$$

It can be shown that the *B* is the optimized lest-square approximation of *A*, and SVD is an orthogonal transformation, there is no relationship between different singular vectors [18]. Obviously, classification with the matrix *B* will faster than classification with matrix *A* because the component of *B* is smaller than *A*. Futhermore, if the data is embedded by noise, one byproduct of dimensionality reduction is noise reduction, the classification accuracy may also be upgraded. An example of the approximation of a time serie with various lower-rank matrix is illustrated in Figure 1.

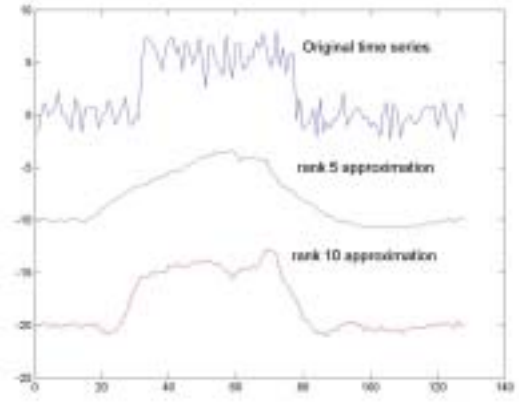

Figure 1. An example of a time series and its low-rank approximation

Considering the matrix *A* containing *n* noise time series $\vec{y}_t = \{y_1, y_2, \ldots, y_d\}$

$$A = \begin{bmatrix} \vec{y}_{t1} \\ \vec{y}_{t2} \\ \vdots \\ \vec{y}_{t3} \end{bmatrix} \quad (6)$$

each time series $\vec{y}_t$ can be represented as a desired time series $\vec{x}_t$ plus an additive noise $\vec{n}_t$, i.e.,

$$\vec{y}_t = \vec{x}_t + \vec{n}_t \quad (7)$$

The basic idea of SVD-based noise reduction algorithms is that some of the singular vectors and their corresponding singular values of *A* contain $\vec{x}_t$, where other singular value or singular vector relate to only noise [20]. Normally, we can assume the amplitude of noise is much lower than that of desired singal, thus the dimensions related to small singular values are viewed as related to noise. The $\vec{x}_t$ can be reconstructed using the information only corresponded with dominate eigenvalues.

However, as pointed out by Ding [21], how to decide the 'small of amplitude' of the singular values is a crucial problem. In addition, for classification problem, the relationship of classification accuracy and dimensionality is nonlinear [1]. It is difficult to choose the threshold for removing noise directly. We propose the wrapper feature extraction addressing this problem by using the classification performance with different lower-rank approximation matrix for deciding which approximation matrix is better.

3. 2. SVD-Based Wrapper Feature Extraction

When classifying the features extracted by SVD, we often assume the dimensions correponding to the large singular values are much important than that corresponding to the small singular values. In the ideal case, the classfication accuracy will monotonically increase when increasing the rank of approximation matrix before meeting with noise related singular values. Howver, for many datasets, the ideal situation will not hold. The classification error with various approximation matrix rank for time series data can be observed in the experimental part. In order to find the rank associated with highest classification accuracy, we borrow the idea from wrapper feature selection to our problem that searching the state space of possible combination of features. The general process of SVD-based wrapper feature extraction is shown in Figure 2.

For the wrapper feature selection algorithm, the possible combination of the features is 2^d for *d* features [2]. Thus it is not feasible for searching the full feature state space exhaustively. Heuristic search algorithms such as Hill-climbing algorithm, best-first searching, and compound searching have been used for wrapper feature selection. For our problem, by using SVD, the features have been ordered by the amplitude of singular values and the singular values are orthogonal to each other. If we take the assumption that large singular values are more important than small singular values, the simple forward selection or backward elimination method is enough for searching the state space. Forward selection refers to the search that starts from the empty set of features, and smoothly increase the dimensionality. Backward elemination begins at the full set of features, and decrease the dimensionality gradually. Thus the candidate searching state is just *r*, the rank of the time series matrix *A*. The algorithm of feedforward selection algorithm for SVD-based wrapper feature extraction is given in Table 1. The backward elemination algorithm for SVD-based wrapper feature extraction is similar to the forward searching algorithm, the only difference is the searching direction. The number of features exceed

a threshold or the estimated classification error lower than a threshold can be used as the stopping criterion [2].

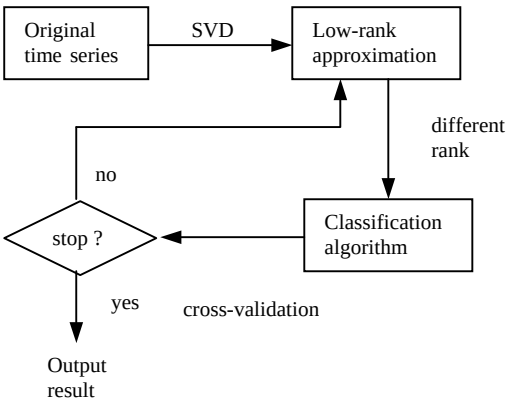

Figure 2. The general process of wrapper feature extraction using SVD

Table 1. The forward search algorithm for SVD-based wrapper feature extraction

Input: The trained time series matrix A_1 and the matrix need to be labeled A_2.

1. Decompose A_1 and A_2 to singular vectors and singular values using SVD.
2. Set the rank *r* equals to 1.
3. Generate the lower-rank approximation matrix B_1 for A_1 and B_2 for A_2 using first few singular values $s_1, s_2, \ldots, s_r$ by Equ. (4).
4. Construct a classifier using B_1.
5. Estimate the classification error on B_2 by cross-validation.
6. If the stop criterion does not satisfied, *r = r+1*, go to step 3.

Although the state space need to be searched is not so huge by using SVD, as we need to perform the classification algorithm for every lower-rank approximation matrix, and normally the classification needs long time to converge, a heuristic that can narrow the searching space will speed the wrapper feature extraction process significantly. As mentioned in section 3.1, if we assume the data is stained by noise, the signal subspace can be separated from the noise subspace with respects to the amplitude of their corresponding singular values. If we can estimate the threshold differentiating the signal subspace from the noise subspace, that is, to determine a threshold γ such that

$$s_r > \gamma > s_{r+1} \tag{8}$$

We may assume that that lower-rank approximation matrix associated with $\{s_1, s_2, \ldots s_r\}$ close to the point has highest classification accuracy. We propose a search method starts from the point *r*, toward the lowest rank and highest rank. We call this search method *bi-direction* search. The difficulty for this method is how to estimate the threshold point *r*. We propose a simple threshold estimation using the variance obtained from a moving sliding window across the singular value sequence. Let

$$\tilde{s}_i = \frac{s_i}{\sum_{i=1}^{r} s_i} \tag{9}$$

be the normalized singular values of the matrix *A*. An example of the normalized singular values with various ranks for a time series data set is shown in Figure. 3.

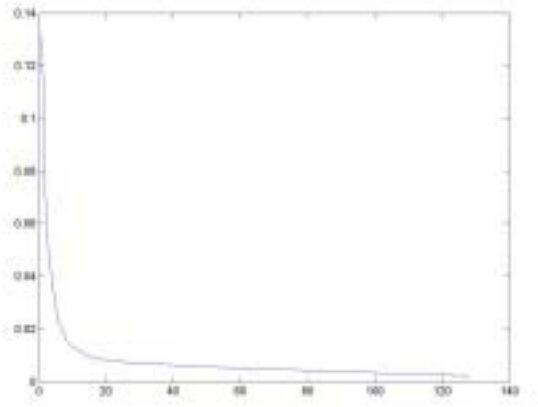

Figure 3. An example of the normalized singular value sequence for a time series data set

As mentioned in Section 3.1, the singular values corresponding to noise are often assumed small compared to that corresponding to the true signal. If we know the character of the noise embedded in time series, we can estimate the threshold by analyzing the statistical property of noise. For example, if we assume the noise is Gaussian noise and we know the standard deviation of the noise σ, Konstantinides and Yao demonstrated that the simple threshold $\gamma = 3\sigma$ is a stable estimation under a variety of noise levels [24]. For our general time series mining task, we don't know the type of noise and the standard deviation is difficult to be estimated. Instead of using the singular value directly, we address the problem of estimation the differentiating point by using the normalized variance of a moving sliding window. The variance of the singular value sequence within a sliding window is defined as

$$\mathrm{var}(\tilde{s}_i) = \frac{\sum_{i=1}^{l}\left(\tilde{s}_i - \bar{\tilde{s}}\right)}{l-1} \quad (10)$$

where l is the length of the sliding window, and $\bar{\tilde{s}}$ is the mean of the normalized singular values within the sliding window. In our experiments, we set the l equals to 3, thus the calculated $\mathrm{var}(\tilde{s}_i)$ will begin from the second singular value and end at the *r-1* singular value. Normally, the first singular value and the last singular value are not the break point between true signal and noise, thus this definition of sliding window should not affect the classification result. For the true signals, the amplitude of singular values are big, and when increasing the dimensionality, the singular value are monotonically decrease, thus the defined variance should be large. Adversely, the defined variance that related to noise will be small. In the interaction point of the single and noise, by assume the amplitude of the true single is much larger that of the noise, the var should have fluctuation. Thus we define the differentiating point as the point corresponding to the variance changes their trend from decreasing to a increasing. Figure 4 illustrates an example of the variance with moving sliding window for a time series dataset.

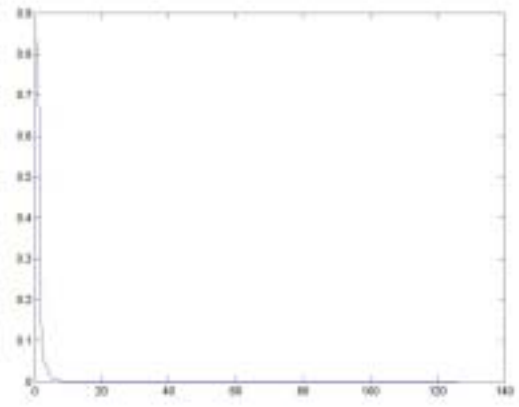

Figure 4. An example of the variance with moving sliding window for a time series data

After calculating the differentiating point *r*, the *bi-direction* search algorithm will move to *r-1* point when the search iteration is an odd number, and search to the *r+1* point when the search iteration an even number.

4. EXPERIMENTAL EVALUATION

We validate the proposed methods by using several benchmark time series data. The main characteristics of the used datasets are presented in section 4.1. The experimental results produced by several searching method are given in Section 4.2. We use two classification algorithms. One is the 1-Nearest Neighbour (1-NN) classification algorithm with the most popular Euclidean distance. Another classification algorithm is the probabilistic RBF neural networks (PNN) [25]. The classification accuracy within the loop is measured by 5-fold cross validation. The final classification accuracy is validated by leave-one validation.

Table 2. The bi-direction search algorithm for SVD-based wrapper feature extraction

Input: The trained time series matrix A_1 and the matrix need to be labeled A_2.

1. Decompose A_1 and A_2 to singular vectors and singular values using SVD.
2. Estimate the differentiating rank *r*.
3. Generate the lower-rank approximation matrix B_1 for A_1 and B_2 for A_2 using first few singular values $s_1, s_2, \ldots, s_r$ by Equ. (4).
4. Construct a classifier using B_1.
5. Estimate the classification error on B_2 by cross-validation.
6. If the stop criterion does not satisfied and the current search iteration is an odd number, *r = r-1*, go to step 3. If the stop criterion does not satisfied and the current search iteration is an even number, *r = r+1*, go to step 3.

4. 1. Data Description

We take four artificial classified time series data sets from UCR time series data mining Archive [22], Another data set is downloaded from Internet. Totally there are six data sets that data labels are available in UCR Archive. The Auslan data set contains multivariate time series, each time series corresponds to several observations, with which we can't apply the proposed algorithm directly. Realitycheck data only has one instance in each class that is suitable for evaluating clustering algorithms not classification algorithm. Therefore, we take the other four data sets in our experiment.

- Cylinder-Bell-Funnel (CBF): It is an artificial data set containing three types of time series: cylinder (c), bell (b), and funnel (f). It has been widely used for validating different data mining algorithms and similarity measures for time series. We generated 128 time series for each class with length 128.
- Control Chart Time Series (CC): This data set has 100 instances for each of six different classes of control charts.

- Trace data set (Trace): The 4 – class data set consisted of 200 instances, 50 for each class. The dimensionality of the data is 275.
- Gun Point data set (Gun): The data set has two classes, each including 100 samples. The dimensionality of the data 150.
- The ECG data set (ECG): This data set was obtained from the ECG database at http://www.physionet.org [23]. We used 3 groups of those ECG time series in our experiments. Group 1 included 22 time series representing the 2 sec ECG recordings of people having malignant ventricular arrhythmia; Group 2 included 18 time series that are 2 sec ECG recordings of healthy people representing the normal sinus rhythm of the heart; Group 3 included 39 time series representing the 2 sec ECG recordings of people having supraventricular arrhythmia.

4. 1. Experimental Results

The classification accuracies using the 1-NN algorithm and Probabilistic Neural Networks PNN for CBF, CC, Trace, Gun and ECG data with various dimensionality is illustrated in Figure 5 and Figure 6, respectively. The results show that when increasing the rank of the approximation matrix values, the classification error will not monotically decrease. Thus wrapper approach is useful for such type of data.

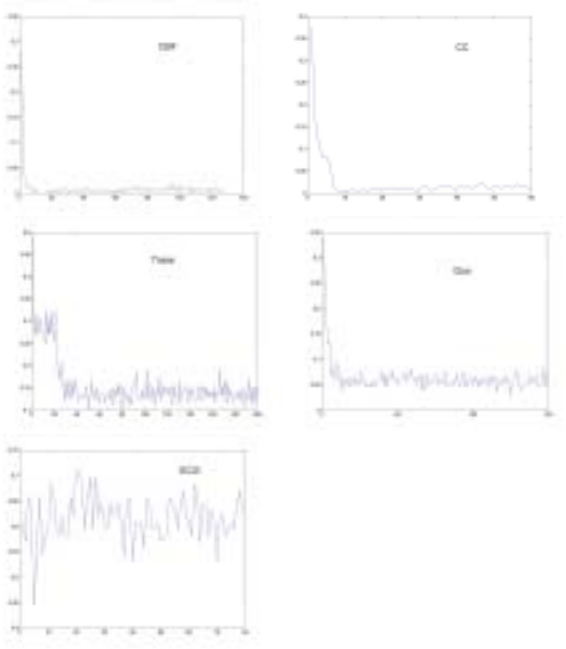

Figure 5. The classification error obtained with various rank for CBF, CC, Trace, Gun and ECG datasets using 1-NN

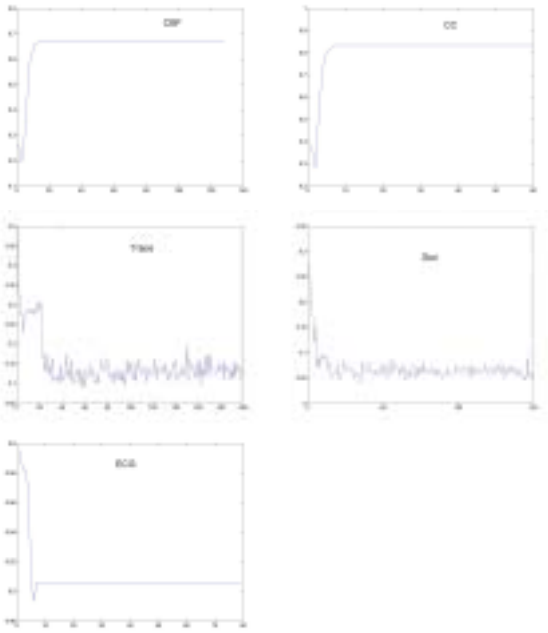

Figure 6. The classification error obtained with various rank for CBF, CC, Trace, Gun and ECG datasets using PNN

The classification error for the original data sets for the used two classifiers is shown in Table 3. Table 4 gives the classification error with the SVD-based wrapper feature extraction algorithm for the used two classifiers. The classification is obtained by leave-one-out cross validation. The wrapper feature extraction algorithm achieves better classification accuracy in three datasets than using the original dataset, and obtains the same classification accuracy in other two datasets with 1-NN algorithm. For PNN classification algorithm, the proposed wrapper feature extraction algorithm produces the same classification error only with Gun dataset, and performs better on other four datasets. Wrapper feature extraction can improve the classification accuracy

Table 3. The classification error (%) for the original data sets

Data set	CBF	CC	Trace	Gun	ECG
1-NN	0.26	1.33	11	5.5	62.03
PNN	66.67	83.33	10	6	50.53

For fair comparison to various searching method, we set the stopping criterion as the classification error estimated by 5-fold cross-validation lower than 0. The searching method that can find the rank with lowest classification error with shortest step is thought better than other methods. The shortest steps for finding the best classification accuracy with various search method using 1-NN and PNN are presented in Table 5 and Table 6, respectively.

Table 4. The classification error (%) gotten by Wrapper Feature Extraction

	CBF	CC	Trace	Gun	ECG
1-NN	0	0.17	11	5.5	45.57
PNN	16.15	25.83	9	6	49.37

Table 5. The shortest search steps with various search methods for 1-NN

	CBF	CC	Trace	Gun	ECG
Forward	10	10	35	18	5
Backward	74	51	76	1	75
Bi-direction	3	6	53	25	1

Table 6. The shortest search steps with various search methods for PNN

	CBF	CC	Trace	Gun	ECG
Forward	3	2	58	35	6
Backward	126	59	107	44	74
Bi-direction	2	2	23	89	1

For 1-NN algorithm, bi-direction search achieves the best classification accuracy with shortest steps in three data sets, forward and backward search method perform best in one data set, respectively. Table 6 demonstrates that bi-direction algorithm can obtain the best classification accuracy with shortest steps in CBF, CC, Trace, Gun and ECG data when using PNN. The forward search performs best on Gun dataset with PNN classification algorithm. The experimental results showed that bi-direction method is better than forward and backward search method globally for the used data sets.

5. CONCLUSIONS

We proposed a wrapper feature extraction approach utilizing the lable information for time series classification. Singular Value Decomposition (SVD) is used as a tool for dimensionality reduction. The proposed wrapper feature extraction using the estimated classification accuracy to select a good lower-rank approximation matrix for the time series data set. We propose three efficient search methods by exploiting the characters of SVD. Experiments performed on several benchmark time series data sets demonstrated that the SVD-based wrapper feature extraction can achieve higher accuracy than using the original data. The bi-direction search method outperforms the forward and backward search methods for the used data sets.

REFERENCES

[1]. R. Bellman , 1961. Adaptive control process: a guided tour, Princeton University Press.

[2]. R. Kohavi and G. H. John, 1997. Wrappers for Feature Subset Selection, Artificial Intelligence 97(1-2): 273-324.

[3]. H. Liu and H. Motoda, 1998. Feature extraction, construction and selection: a data mining perspective, Kluwer Academic Publishers.

[4]. K. Kanth, D. Agrawal, A. Abbadi, A. Singh. 1998. Dimensionality reduction for similarity searching in dynamic databases. In Proceedings of the 1998 ACM SIGMOD International Conference on Manangement of Data, pp. 166-176

[5]. S. Chandrasekaran, B. S. Manjunath, Y. F. Wang, J. Winkeler, H. Zhang, 1997. An eigenspace update algorithm for image analysis. CVGIP: Graphical Models and Image Processing Journal, 59(5): 321-332.

[6]. V. Castelli, A. Thomasian, C. S. Li, 2003. CSVD: Clusterng and singular value decomposition for approximate similarity search via high-dimensional spaces. IEEE Trans. on Knowledge and Data Engineering, 15(3): 671-685.

[7]. F. Korn, H. Jagadish, C. Faloutsos, 1997. Efficiently supporting ad hoc queries in large datasets of time sequences. In Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data. pp. 289-300.

[8]. L. Guyon and A. Elisseeff, 2003. An introduction to variable and feature selection. Journal of Machine Learning Research, 3:1157-1182.

[9]. S. Petridis, S. J. Perantonis, 2004. On the relation between discriminant analysis and mutual information for supervised linear feature extraction. Pattern Recognition, 37(5): 857-874.

[10]. K. Kira and L. Rendell, 1992. A practical approach to feature selection. In Proceedings of the Ninth International Conference on Machine Learning, pp. 249-256.

[11]. I. Kononenko, 1994. Estimating attributes: analysis and extensions of relief. In Proceedings of the 1994 European Conference of MachineLearning, pp. 171-182.

[12]. M. A. Hall, 2000. Correlation-based feature selection for discrete and numeric class machine learning. In Proceedings of the 17th International Conference on Machine Learning, pp. 359-366.

[13]. H. Peng, F. Long, and C. Ding, 2005. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. on Pattern Analysis and Machine Intelligence, 27(8): 1226-1238.

[14]. L. Yu and H. Liu, 2004. Efficient feature selection via analysis of relevance and redundancy, Journal of Machine Learning Research, 5: 1205-1224.

[15]. C. Ding and H. C. Peng, 2003. Minimum redundancy feature selection from microarray gene expression data. In Proceedigns of the Second IEEE Computational Systems Bioinformatics Conference, pp. 523-528.

[16]. R. Agrawal, C. Faloutsos, and A. Swami, 1993. Efficient similarity search in sequence databases. In Proceedings of the 4th Conference on Foundations of Data Organization and Algorithms, pp. 69-84.

[17]. K. P. Chan, A. W. Fu, and T. Y. Clement, 2003. Haar wavelets for efficient similarity search of time series: with and without time warping. IEEE Trans. Knowledge and Data Engineering 15(3): 686-705.

[18]. D. Kalman, 1996. A singularly valuable decomposition: The SVD of a Matrix, The College Mathmatics Journal, 27: 2-23.

[19]. T. K. Landauer and S. T. Dumais, 1997. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychological Review, 104(2): 211-240.

[20]. M. Dendrinos, S. Bakamidis, and G. Carayannis, 1991. Speech enhancement from noise: A regenerative approach. Speech Communication, 10: 45-57.

[21]. C. Ding, 1999. A similarity-based probability model for latent sementic indexing. In Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.

[22]. E. Keogh and T. Folias, 2002. The UCR Time Series Data Mining Archive, http://www.cs.ucr.edu/~eamonn/TSDMA/index.html.

[23]. A. L. Goldberger, L. A. N. Amaral and L. Glass and J. M. Hausdroff and P. ch. Ivanov and R. G. Mark and J. E. Mietus and G. B. Moody and C. –K. Peng and H. E. Stanley, 2000. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulalation, 101(23): e215-e220.

[24]. K. Konstantinides, K. Yao, 1988. Statistical analysis of effective singular values in matrix rank determination. IEEE Tran. Acoust. Speech, Signal Process. 36: 757-763.

[25]. D. F. Specht, 1990. Probabilistic Neural Networks, Neural Networks, 11(3): 109-118