Time Warping Symbolic Aggregation Approximation with Bag-of-Patterns Representation for Time Series Classification

Zhiguang Wang
Department of Computer Science
and Electric Engineering
University of Maryland Baltimore County
Email: stephen.wang@umbc.edu

Tim Oates
Department of Computer Science
and Electric Engineering
University of Maryland Baltimore County
Email: oates@umbc.edu

***Abstract*—Standard Symbolic Aggregation approXimation (SAX) is at the core of many effective time series data mining algorithms. Its combination with Bag-of-Patterns (BoP) has become the standard approach with state-of-the-art performance on standard datasets. However, standard SAX with the BoP representation might neglect internal temporal correlation embedded in the raw data. In this paper, we proposed time warping SAX, which extends the standard SAX with time delay embedding vector approaches to account for temporal correlations. We test time warping SAX with the BoP representation on 12 benchmark datasets from the UCR Time Series Classification/Clustering Collection. On 9 datasets, time warping SAX overtakes the state-of-the-art performance of the standard SAX. To validate our methods in real world applications, a new dataset of vital signs data collected from patients who may require blood transfusion in the next 6 hours was tested. All the results demonstrate that, by considering the temporal internal correlation, time warping SAX combined with BoP improves classification performance.**

I. Introduction

Time series data is ubiquitous in many real world areas, such as health care, finance, geography, information technology, etc. However, it suffers from high dimensionality and noise. As one type of the successful techniques to discretize and reformulate raw time series data, symbolic time series analysis has been used in many different application areas to identify temporal patterns [1]. Aligned Cluster Analysis (ACA) was introduced as an unsupervised approach to cluster the temporal patterns on the human motion data [2]. It is an extension of kernel k-means clustering but requires significant computational capacity. Persist is an unsupervised discretization method to maximize the persistence measurement of each symbol [3]. The Piecewise Aggregation Approximation (PAA) method was proposed by Keogh [4], which then upgrades to Symbolic Aggregation approXimation (SAX) [5]. In SAX, each aggregation value after the PAA process is mapped into equiprobable intervals according to a standard normal distribution to produce symbolic representations. SAX has become a common representation method due to its simplicity and effectiveness on various types of data mining tasks [6].

SAX forms the PAA in temporal order using a sliding window. Such representation is effective in several data mining tasks such as indexing [6] and visualization [7]. As one of the effective features for classification, a bag-of-words makes use of SAX words to encode non-linearity and benefits from invariance to rotations [8]. Lin *et al*. also reported the state-of-the-art results using a One-Nearest-Neighbor classifier (1NN) on UCR Time Series Classification/Clustering databases [9]. Oates *et al*. applied SAX and BoP to predict outcomes for traumatic brain injury patients [10] and explored representation diversity by ensemble voting to further improve classification performance [11].

While standard SAX with BoP obtains the state-of-the-art performance on benchmark datasets and real world applications, we hypothesize that if we could improve SAX by capturing more temporal information in the BoP representation. Our work is inspired by two observations. First, correlation is common in time series. Statistical time series analysis utilizes AutoCorrelation Function (*ACF*) and Partial AutoCorrelation Function (*PACF*) to interpret the internal linear correlations in ARIMA models [12]. We propose to explicitly extract the implicit linear correlation in time series and thus to help reveal the intrinsic statistical properties to potentially improve the expressive capacity.

Another motivation is based on the observation that BoP extracts local patterns, regardless of where they occur within time series. Standard SAX keeps the temporal ordering information while this information might be ignored after building a BoP. We attempt to seal the temporal correlation in the bags by borrowing the concept of time embedding vector from dynamic systems to overcome the loss of information. In this paper, our contributions are:

- We extend standard SAX with a time warping procedure on three granularity levels, i.e. items, bins and windows, to capture the temporal information within the time series.
- After building a BoP from time warping SAX words, we compare its classification performance with standard SAX on 12 benchmark datasets and a new real world dataset from patient vital signs. Our experiments justify our assumption that capturing temporal correlation within SAX words and

BoP helps improve the expressive capacity and improve classification accuracy.

II. MOTIVATION

A. Internal Linear Correlations Embedded in Time Series

For a stationary process $Z = Z_1, Z_2, ..., Z_t$, the covariance between Z_t and Z_{t+k} is defined as:

$$\gamma_k = cov(Z_t, Z_{t+k}) = E[(Z_t - \mu)(Z_{t+k} - \mu)] \quad (1)$$

The correlation between Z_t and Z_{t+k} is:

$$\rho_k = \frac{cov(Z_t, Z_{t+k})}{\sqrt{var(Z_t)}\sqrt{var(Z_{t+k})}} \quad (2)$$

As a function of k, ρ_k is called the AutoCorrelation Function (*ACF*). It represents the correlation between Z_t and Z_{t+k} at time lag k.

If we remove the mutual linear dependency on the intervening variables $Z_{t+1}, Z_{t+2}, ..., Z_{t+k-1}$, the conditional correlation is given by:

$$corr(Z_t, Z_{t+k}|Z_{t+1}, Z_{t+2}, ..., Z_{t+k-1}) \quad (3)$$

The resulting value is called the Partial AutoCorrelation Function (*PACF*) in time series analysis.

ACF and *PACF* helps discover the internal correlation and identify the order of ARIMA model in statistical time series analysis [12]. Time series have intrinsic correlations even though sometimes we observe no obvious periodic trends. As shown in Figure 1, although there is no obvious periodic trend in the raw data, the *ACF* and *PACF* values show significant linear dependencies at different time lags. In this example, the raw data needs a first order or even higher order differencing to reveal the essential linear temporal dependency in a stationary time series.

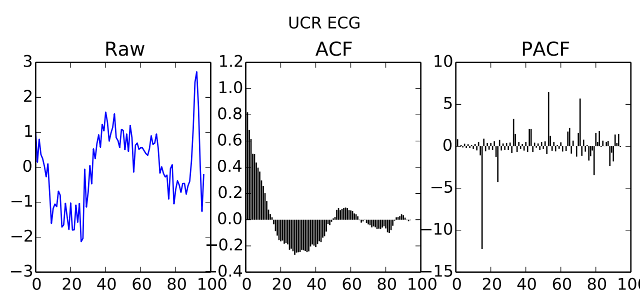

Fig. 1. Plots of (a) raw data (left), (b) *ACF* (middle) and (c) *PACF* (right) on ECG dataset from UCR Time Series Classification/Clustering database.

Correlation revealed by *ACF* and *PACF* are more general than periodicity and is very commonly observed in time series data. A number of *ACF* and *PACF* plots from UCR time series databases with various type of data show similar phenomena as seen in Figure 1.

In Figure 2, standard SAX word shows their intrinsic property of preserving the major internal correlation embedded in raw data. This is because standard SAX builds PAA bins and slides windows sequentially in temporal order.Sequential

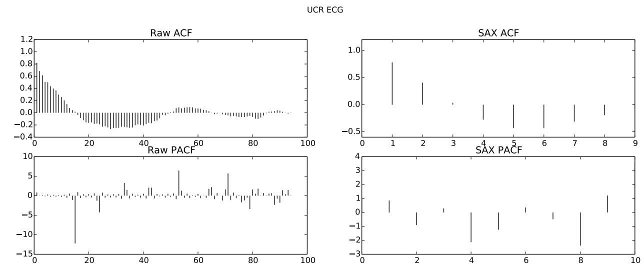

Fig. 2. Plots of (a) raw data *ACF* (left up), (b) SAX word *ACF* (right up) and (c) raw data *PACF* (left bottom) and (d) SAX word *PACF* (right bottom)on ECG datasets from UCR Time Series Classification/Clustering database. Standard SAX preserves the temporal correlation, but this advantage might be ignored in BoP representations.

information in temporal order is one of the essential properties of time series, standard SAX works quite well on tasks like similarity evaluation and time series indexing. However, this advantage might be lost in the BoP representation. We discuss this in more detail in the next section.

B. Order Invariance in BoP Representation

A Bag-of-Patterns (BoP) is a histogram-based representation for time series data, similar to the bag-of-words approach that is widely accepted by the natural language processing and information retrieval communities. Given a time series of length L, a BoP representation is constructed by sliding a window of length n ($n << L$) to map each subsequence of length n into a SAX word. Then the $L - n + 1$ subsequence is represented as a histogram of word counts.

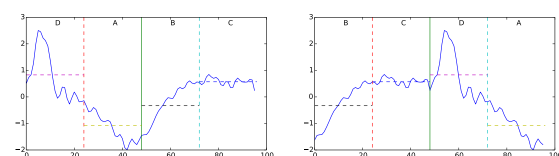

Fig. 3. PAA and SAX procedure for ECG data. (a) Time series are partitioned into two windows and each window has two PAA bins (left). We exchange the order of two windows in (b) (right), but the BoP patterns are same, i.e, *BC*: 1, *DA*: 1

BoP can handle time series with varying length. It is invariant to the shift of pattern locations by extracting local structures regardless of where they occur. Nonetheless, the standard SAX approach preserves the temporal order by sliding windows sequentially. BoP runs the risk of losing this temporal dependency due to its shift invariance. To clearly state the problem, consider the toy example as shown in Figure 3. We partition the time series into two windows with two PAA bins in each half (Figure 3(a)). Then we exchange the order of the two windows as shown in Figure 3(b). Obviously, these two time series have different temporal information from each other, but they produce the same BoP.

The BoP reveals the higher level structures, its invariance to shifts also allows rotation-invariant matching in shape datasets. In the next section, we introduce time warping SAX approaches. They capture the temporal correlation information embedded in time series by taking advantage of the invariance to shift and temporal order in BoP to generate more informative representations.

III. TIME WARPING SAX

In this section, we introduce three time warping SAX approaches inspired by the idea of time delay embedding vectors [13], [14]. They take advantage of the order-invariance of BoP to capture temporal correlation information through building words and bags with a time warping procedure.

Given a time series $T = \{t_1, t_2, ..., t_L\}$ of length L, a sliding window of length n and a number of PAA w, the standard SAX method partitions the time series into $\lceil \frac{L}{n} \rceil - 1$ equal-sized sliding windows $t_i, t_{i+1}, ..., t_{i+n-1}$. Each window is then divided into w PAA bins. In time warping SAX, we change the step size $i = i + 1$ by $i = i + \tau$ to build the new time warping sequence $t_i, t_{i+\tau}, ..., t_{i+\tau(n-1)}$. The delay-time embedding vector $t_i, t_{i+\tau}, ..., t_{i+\tau(n-1)}$ is used to embed the time series into a higher dimensional space to reconstruct the original state space vector. τ is the delay time and n is the embedding dimension. In the time warping procedure, τ has the same mechanism with the time lag k in *ACF* and *PACF*. The embedding dimension n is equivalent to the sliding window size.

In this section, we will introduce two time warping SAX approaches named *Skipword* and *Skipbin*. They correspond to the different discretization granularity in standard SAX. Moreover, we explore the effect of a skip step on sliding window in standard SAX and call it the *Skipwindow* SAX approach.

A. Skipword SAX

Skipword SAX applies the time warping approach to build PAA bins. Because each corresponding SAX word is the mean value of PAA bins, the idea of *Skipword* is to seal the temporal correlations into single SAX word through delay-time embedding vectors.

Consider a simple sequence $T = \{1, 2, ..., 10\}$ with delay time $\tau = 2$, embedding dimension $n = 6$ and number of bins $w = 3$. T is divided into three windows with three PAA bins in each window as $\{1\ 3 \mid 2\ 4 \mid 3\ 5\}$, $\{6\ 8 \mid 7\ 9 \mid 8\ 10\}$. After calculating the mean values in the PAA bins, T is discretized as $\{2\ 3\ 4\}$, $\{7\ 8\ 9\}$. For simplicity, we skip the Gaussian mapping procedure in the following examples to map the rounding of each PAA mean to its corresponding SAX word with the dictionary {1: A, 2: B, ..., 26: Z}. Thus, we get the BoP pattern *BCD*: 1, *GHI*: 1. Then the window slides forward to generate more BoP patterns with the same loop. The over-simplification here is only to facilitate the explanation of our approaches. In our experiments, we process the data with the full pipeline of standard SAX including the Gaussian mapping.

B. Skipbin SAX

Skipbin SAX applies time warping approaches to build the sliding window instead of PAA bins. Each window is cut into PAA bins following the standard SAX procedure. By enlarging the granularity of temporal correlations, we expect different embedding information to be sealed in BoP representations.

Consider the same sequence $T = \{1, 2, ..., 10\}$ with the delay time $\tau = 2$, embedding dimension $n = 5$ and the number of bins $w = 3$. We apply time warping on subsequence to build two windows and partition each window into three PAA bins $\{1\ 3 \mid 5\ 7 \mid 9\}$, $\{2\ 4 \mid 6\ 8 \mid 10\}$. After figuring out the mean values in each PAA bins, T is discretized into $\{2\ 6\ 9\}$ and $\{3\ 7\ 10\}$ with the BoP patterns *BFI*: 1, *CGJ*: 1. Then The window slides forward to generate more BoP patterns in the same way.

C. Skipwindow SAX

Skipwindow SAX approach is the same as standard SAX. The only difference is the changing time steps among each sliding windows. For the sequence $T = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, we set time step $t = 5$, window length $n = 5$ and words number $w = 3$. The *Skipwindow* SAX discretizes the sequence into $\{1\ 2 \mid 3\ 4 \mid 5\}$, $\{6\ 7 \mid 8\ 9 \mid 10\}$. No overlaps occurs in this example due to the large skip size ($t = 5$). The final BoP patterns are *BDE*: 1, *GIJ*: 1.

D. Summarization and Discussion

By using time warping procedure based on delay-time embedding vectors, time warping SAX approaches seal the correlation information into different sizes of "bags", i.e., PAA bins and sliding windows. *Skipword* has smaller granularity because each SAX word will contain temporal correlation with different delay embedding vector parameters. Instead of embedding the correlation into each single SAX word, the *Skipbin* approach looks at larger temporal scales to encode the correlation into the windows that contain several PAA bins. *Skipbin* SAX words have longer memory than *Skipword* SAX words because the temporal order is preserved in sliding windows with larger scale than a single PAA bin. *Skipwindow* SAX is a straight-forward extension of the standard SAX approach with a larger parameter space as the time step t will change instead of fixing it at 1. We assume that *Skipwindow* also captures some temporal dependency when building the BoP representations. However, when the skip size increases, the BoP discards more phases of the original time series and results in more information loss.

Delay-time embedding is a powerful tool. In principle, almost any delay time τ and embedding dimension n works when we need to analyze the correlation behavior of a complex dynamical system if we have unlimited precise data. However, choosing the embedding vector $\tau(n-1)$ is not trivial. We want both τ and and $\tau(n-1)$ to be close to some characteristic decorrelation time. One suggested principle is exactly the *ACF* [15].

Note that if we tune the parameters of the three time warping SAX approaches, the standard SAX has a chance to be replicated. That is, standard SAX is a specific subset of time warping SAX. While we take temporal correlation into consideration, we also add one more dimension, the time delay τ in the parameter space. To avoid the "cheat" situation and reveal the real impact of the internal correlation, we get rid of the parameter intervals that will reduce time warping SAX to the standard SAX.

IV. EXPERIMENTS AND RESULTS

A. Data

The benchmark datasets are from the UCR Time Series Classification and Clustering home page [9]. We choose the subsets for which the BoP error rate with standard SAX is above 0.1. The error rate is the fraction of incorrectly classified instances. For each dataset, the table below gives its name, the number of classes and the length of the individual time series. The datasets are pre-split into training and testing sets to facilitate experimental comparisons. We also give the number of training and test instances. The test error rate of the standard SAX-BoP approach is directly taken from [11].

TABLE I
BENCHMARK TIME SERIES DATASETS AND SUMMARY STATISTICS

	Class	Train	Test	Length	Error Rate [11]
Coffee	2	28	28	286	0.1071
Oliveoil	4	30	30	570	0.1667
ECG200	2	100	100	96	0.22
Lighting2	2	60	61	637	0.2295
Lighting7	7	70	73	319	0.3973
Beef	5	30	30	470	0.4667
Adiac	37	390	391	176	0.3836
50words	50	450	455	270	0.4396
FaceAll	14	560	1690	131	0.2497
OSULeaf	6	200	242	427	0.3058
SwedishLeaf	15	500	625	128	0.2064
yoga	2	300	3000	426	0.1677

The second dataset is the patient vital signs signals (ECG and PPG) collected from University of Maryland School of Medicine. All patient data are anonymous in order to protect patient privacy. 556 patient's ECG and PPG data were collected in 68 to 128 minutes at a 240 Hz sampling rate. Among them, 237 patient's vital signs data are less than 128 minutes long and all the data is quite noisy. The label demonstrates if the patient needed a blood transfusion for Packed Red Blood Cell (pRBC) within 6 hours of admission. The vital signs time series are preprocessed by filtering outliers and integrating the means in each minute to reduce data size. Because the data is highly skewed with only 17 positive instance, we down-sampled 17 negative instance to rebuild a balanced dataset.

B. Experiment and Analysis

We construct BoPs for the datasets in Table I by looping over the hyperparameters on the training set. Given a time series of length m, we set $n \in \{0.15m, 0.16m, ..., 0.36m\}$,$w \in \{2, 4, 6, 8\}$ and $a \in \{3, 4, ..., 10\}$. Moreover, the delay-time τ and skip time t loops in $\{0.01m, 0.03m, ..., 0.15m\}$. We classify the time series with each BoP on the training set to select the optimal parameters with Leave-One-Out Cross Validation (LOOCV) for the test set. If two or more representations tie, we choose the representation with the smallest possible vocabulary size a^w and delay-time τ. For vital signs data, we apply LOOCV to evaluate the classification performance on the training set alone.

TABLE II
1NN ERROR RATE OF TIME WARPING SAX AND STANDARD SAX ON TEST DATASETS

Dataset	SAX	*Skipword*	*Skipbin*	*Skipwindow*
Coffee	0.1071	**0.0357**	**0.0357**	**0.0357**
Oliveoil	0.1667	**0.1**	**0.1**	**0.1**
ECG200	0.22	**0.12**	**0.12**	**0.12**
Lighting2	0.2295	0.2295	**0.1475**	**0.1475**
Lighting7	0.3973	0.2603	**0.2466**	0.3288
Beef	0.4667	0.4333	**0.4**	0.4667
Adiac	**0.3836**	0.5166	0.5141	0.5166
50words	0.4396	0.3867	**0.3756**	0.3978
FaceAll	0.2497	0.2438	0.2438	**0.2349**
OSULeaf	**0.3058**	0.3430	0.3347	0.3554
SwedishLeaf	**0.2064**	0.2400	0.2336	0.2496
yoga	0.1677	0.1670	**0.1603**	0.2017

Table II shows the LOOCV error rate on test sets. Except for "Adiac", "OSULeaf" and "SwedishLeaf", *Skipbin* and *Skipword* SAX methods outperform standard SAX. *Skipwindow* SAX also demonstrates specific improvements, although it is always worse than the *Skipbin* and *Skipword* approaches. Because *Skipwindow* is the direct extension of standard SAX with a larger parameter space, the improvement seems "obvious" as it searches four parameters as opposed to three. However, *Skipwindow* will discard more original data when time step grow large (particularly when $t > w$) and leads to significant information loss.

Among these three time warping SAX approaches, *Skipbin* shows better expressive capacity than the other two based on the classification performance. This is most likely because *Skipbin* SAX applies the time warping procedure to build windows that contains several PAA bins. It takes advantage of delay-time embedding vectors in capturing the temporal correlation in the larger window but also preserves sequential order of the subsequence within each PAA bin. With appropriate combinations of hyperparameters, *Skipbin* extracts the temporal dependency as well as the sequential information to learn more powerful BoP representations.

Fig. 4. Illustration of *Skipword* (left) and *Skipbin* (right) SAX. *Skipword* extracts the temporal correlation through using delay-time embedding vector in each single word; *Skipbin* captures the linear correlation in each word and preserves the sequential order in the PAA bins.

It is necessary to explore the representation diversity between time warping SAX and standard SAX approaches. Different datasets will require different parameters to properly

TABLE III
BEST FOUR REPRESENTATIONS $(n, w, a, \tau/t)$ BY ERROR RATE FOR TWO DATASETS.

Dataset	SAX	*Skipword*	*Skipbin*	*Skipwindow*
50words	94, 8, 3 , 1	65, 4, 3, 27	65, 4, 3, 27	95, 4, 3, 3
	97, 4, 3 , 1	95, 4, 3, 19	68, 4, 3, 27	86, 4, 3, 3
	94, 4, 4 , 1	78, 4, 3, 19	86, 4, 3, 3	90, 4, 3, 3
	86, 8, 3 , 1	89, 4, 3, 19	62, 4, 3, 27	97, 4, 3, 3
Lighting7	70, 8, 4 , 1	48, 2, 8, 3	48, 2, 10, 3	48, 2, 10, 3
	70, 8, 3 ,1	51, 4, 5, 3	73, 2, 10, 41	57, 2, 10, 3
	51, 2, 7, 1	90, 4, 7, 3	57, 2, 10, 22	70, 2, 7, 3
	51, 2, 9, 1	54, 2, 6, 3	60, 2, 10, 41	96, 2, 5, 3

demonstrate the expressive capacity and maximize the classification accuracy [11]. The different time warping procedures in time warping SAX approaches result in the diverse BoP patterns as illustrated in Figure 4.

A time warping SAX representation is determined by the window size n, the number of words w, the alphabet size a and the delay-time/skip step τ/t. Because various datasets and different time warping approaches need specific combinations of hyperparameters to learn appropriate BoPs, one would expect to observe significant representation diversity on different SAX approaches and datasets. We also assume that we can find some patterns in the representation diversities. Table III shows four optimal representations with the best training error rate on two datasets. The best representations for the "50Words" have the same word and alphabet sizes $(4, 3)$ on all time warping SAX approaches. The Window length and delay-time change respectively to learn the BoPs with significant temporal correlations. The "Lighting7" dataset shows more obvious diversity in the best representations, but *Skipbin* SAX approach has the same word and alphabet sizes $(2, 10)$, which means it prefers short words and high alphabetical resolution to build the BoP patterns. On these two datasets, the skip step for *Skipwindow* SAX keeps on the lowest value $(0.01m)$, which is in accordance with our analysis that large skip step brings much information loss. Thus *Skipwindow* SAX is trying to avoid such information loss using small sliding steps.

Figure 5 shows the plots of training error rate on the "50word" and "Lighting7" datasets with all representations sorted in a ascending order over three time warping SAX approaches. Although we observe slight difference among the error rates of best representations for three time warping SAX approaches in Table II, the error rate curve over all representations show more details about the difference among time warping approaches. Three time warping SAX have the similar best performance, but the curve of *Skipbin* SAX stabilizes at the lower horizon with more representations than *Skipword* and *Skipwindow*. It implies that *Skipbin* SAX is muck likely to have robust and better classification accuracy. On "Lighting7", *Skipword* SAX stays at the best error rate at the very beginning but follows by a sharp slope where the classification accuracy rapidly decreases. Through this sharp rising on error rate, the curve leaps to the stable state where the representations show significantly worse classification performance. This means

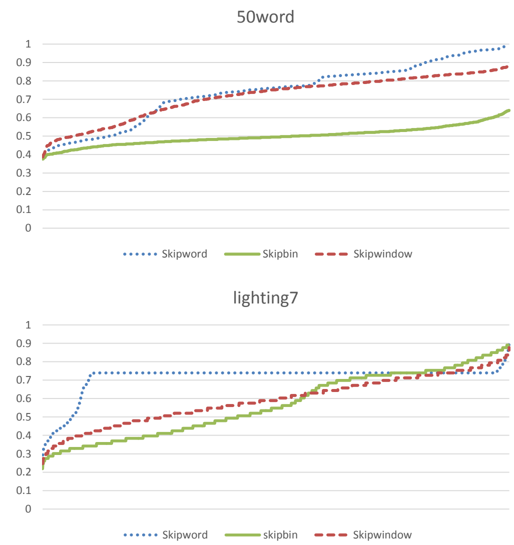

Fig. 5. Curve of test error rate for all representations of time warping SAX by rank (the lowest error is 1) for the "50word" and "Lighting7" datasets.

TABLE IV
ERROR RATE PREDICTING PATIENT OUTCOMES USING VITAL SIGNS WITH DIFFERENT SAX APPROACHES

	ECG	PPG
Standard SAX	0.2353	0.2059
Skipbin SAX	0.2059	0.1765
Skipword SAX	0.2353	0.1765
Skipwindow SAX	0.1765	0.2353

that on some dataset, *Skipword* achieves good classification performance but has a large variance and more significant risk of overfitting. *Skipwindow* SAX always achieve medium performance but with good robustness because the sloping region for both datasets are smooth.

The next experiment explores the utility of time warping SAX to predicting if the patient needs blood transfusion based on their vital sign data. We use 1NN classifier and report their best LOOCV error rate in Table IV.

On ECG dataset, *Skipword* is equivalent to standard SAX while *Skipbin* and *Skipwindow* SAX outperforms standard SAX. On PPG data, *Skipwindow* approach is worse than standard SAX, but other two approaches both overtake standard SAX with 82.35% classification accuracy. These results on real world physiological data support our analysis that *Skipbin* and *Skipword* SAX are more likely to have better classification performance than standard SAX as they capture the temporal correlation and take advantage of the shift-invariance of BoP. *Skipwindow* approach is a natural extension based on standard SAX, its performance might "cheatly" outperform its prototype if we can find the optimal parameters. Time warping SAX has one more parameter (τ or t) which requires more effort to tune the parameters. However, our approaches and results make a strong suggestion to consider the linear or even non-linear temporal correlations when building SAX words.

V. RELATED WORK

Since proposed in 2003, SAX and its derivatives have been successfully applied in many different application areas to identify temporal patterns. To mention a few, Koegh *et al*. introduce the new problem of finding time series discords and apply a SAX derivations to find the subsequences of a longer time series that are maximally different to all the rest of the time series subsequences [16]. Yankov *et al*. introduce a new SAX-based algorithm to discover time series motifs which are invariant to uniform scaling. They show that it produces objectively superior results in several important domains [17]. Vector Space Model is combined with SAX (SAX-VSM)to discover time series patterns and helps classification [18]. They report the state-of-the-art classification performance on UCR standard time series data set. In short, the SAX method has become one of the de facto standard to discretize time series and is at the core of many effective classification or indexing algorithms.

Nonetheless, among a number of papers that explore the application of SAX and its derivatives, no work has been investigated to integrate BoP patterns and the SAX words with linear temporal correlations. The time warping approach to build SAX words in this paper is based on a few research work about symbolic dynamics of time series. The research on permutation entropy [13], [14] and fractal dimensions [15] include details of delay-time embedding vectors. Oates *et al*. apply SAX with bag-of-words to detect brain traumatic injury patients [10] and explore the representation diversity via ensembles of different BoP representations [11] also provide us the analysis paradigm and clear logic map to analyze and apply our approaches on physiological data.

VI. CONCLUSION

Motivated by the internal correlation embedded in time series and intrinsic property of BoP representations, we proposed the time warping SAX to integrate the temporal correlation when building SAX words and BoP representations. Experimental results on 12 standard datasets showed the overall improvements of the classification accuracy on standard SAX-BoP approaches. We analyze the representation diversity of different time warping SAX approaches. The Additional experiments showed that time warping SAX approaches lead to more accurate predictions of patient outcomes based on high resolution vital signs data. Empirically, *Skipbin* SAX has more robust and optimal classification performance as it has less information loss than *Skipwindow* SAX and contains more sequential information than *Skipword* SAX.

Future work will include exploring impact of nonlinear correlation to build SAX word. Except for the internal correlation within univariate time series, how to integrate the cross correlation multivariate time series is also a worth direction. In addition, choosing hyperparameters for time warping SAX algorithm to obtain optimal representation with good generalization capacity is more time consuming than standard SAX. We propose to utilize the heuristic search method to address this issue.

REFERENCES

[1] C. S. Daw, C. E. A. Finney, and E. R. Tracy, "A review of symbolic analysis of experimental data," *Review of Scientific Instruments*, vol. 74, no. 2, pp. 915–930, 2003.

[2] F. Zhou, F. Torre, and J. K. Hodgins, "Aligned cluster analysis for temporal segmentation of human motion," in *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*. IEEE, 2008, pp. 1–7.

[3] F. Mörchen and A. Ultsch, "Finding persisting states for knowledge discovery in time series," in *From Data and Information Analysis to Knowledge Engineering*. Springer, 2006, pp. 278–285.

[4] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, "Dimensionality reduction for fast similarity search in large time series databases," *Knowledge and information Systems*, vol. 3, no. 3, pp. 263–286, 2001.

[5] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*. ACM, 2003, pp. 2–11.

[6] A. Camerra, T. Palpanas, J. Shieh, and E. Keogh, "isax 2.0: Indexing and mining one billion time series," in *Data Mining (ICDM), 2010 IEEE 10th International Conference on*. IEEE, 2010, pp. 58–67.

[7] N. Kumar, V. N. Lolla, E. J. Keogh, S. Lonardi, and C. A. Ratanamahatana, "Time-series bitmaps: a practical visualization tool for working with large time series databases." in *SDM*. SIAM, 2005, pp. 531–535.

[8] J. Lin, R. Khade, and Y. Li, "Rotation-invariant similarity in time series using bag-of-patterns representation," *Journal of Intelligent Information Systems*, vol. 39, no. 2, pp. 287–315, 2012.

[9] E. Keogh, X. Xi, L. Wei, and C. A. Ratanamahatana, "The ucr time series classification/clustering homepage," *URL= http://www. cs. ucr. edu/˜ eamonn/time_series_data*, 2006.

[10] T. Oates, C. F. Mackenzie, L. G. Stansbury, B. Aarabi, D. M. Stein, and P. F. Hu, "Predicting patient outcomes from a few hours of high resolution vital signs data," in *Machine Learning and Applications (ICMLA), 2012 11th International Conference on*, vol. 2. IEEE, 2012, pp. 192–197.

[11] T. Oates, C. F. Mackenzie, D. M. Stein, L. G. Stansbury, J. Dubose, B. Aarabi, and P. F. Hu, "Exploiting representational diversity for time series classification," in *Machine Learning and Applications (ICMLA), 2012 11th International Conference on*, vol. 2. IEEE, 2012, pp. 538–544.

[12] G. E. Box, G. M. Jenkins, and G. C. Reinsel, *Time series analysis: forecasting and control*. John Wiley & Sons, 2013.

[13] C. Bandt and B. Pompe, "Permutation entropy: a natural complexity measure for time series," *Physical Review Letters*, vol. 88, no. 17, p. 174102, 2002.

[14] M. Riedl, A. Müller, and N. Wessel, "Practical considerations of permutation entropy," *The European Physical Journal Special Topics*, vol. 222, no. 2, pp. 249–262, 2013.

[15] S. P. Clark, "Estimating the fractal dimension of chaotic time series," *Lincoln Laboratory Journal*, vol. 3, no. 1, 1990.

[16] E. Keogh, J. Lin, and A. Fu, "Hot sax: Efficiently finding the most unusual time series subsequence," in *Data mining, fifth IEEE international conference on*. IEEE, 2005, pp. 8–pp.

[17] D. Yankov, E. Keogh, J. Medina, B. Chiu, and V. Zordan, "Detecting time series motifs under uniform scaling," in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2007, pp. 844–853.

[18] P. Senin and S. Malinchik, "Sax-vsm: Interpretable time series classification using sax and vector space model," in *Data Mining (ICDM), 2013 IEEE 13th International Conference on*. IEEE, 2013, pp. 1175–1180.