

Pooling SAX-BoP Approaches with Boosting to Classify Multivariate Synchronous Physiological Time Series Data

Zhiguang Wang, Tim Oates

Department of Computer Science and Electrical Engineering
University of Maryland Baltimore County
{stephen.wang, oates}@umbc.edu

Abstract

As the current standard practice of manually recorded vital signs through a few hours is giving way to continuous, automated measurement of high resolution vital signs, it brings a tremendous opportunity to predict patient outcomes and help to improve the early care. However, making predictions in an effective way is fairly challenging, because high resolution vital signs data are multivariate, massive and noisy. Inspired by the max-pooling approaches in Convolutional Neural Networks (CNN), we propose extensions of vanilla SAX-BoP approach, called Pooling SAX-BoP to successfully predict patient outcomes from multivariate synchronous vital signs data. Our experiments on two standard datasets demonstrate the Pooling SAX-BoP approaches are competitive with the current state-of-the-arts on multivariate time series classification problems. We also integrate Boosting algorithm as one of the most powerful ensemble learning approaches on the BoP representations to further improve the performance. Our experimental results on the clinical data demonstrate that our methods are accurate and stable for classifying multivariate synchronous vital signs time series data.

Introduction

Non-invasive, continuous, high resolution vital signs data, such as Electrocardiography (ECG) and Photoplethysmograph (PPG), are commonly used in hospital settings for better monitoring of patient outcomes to optimize early care. Such intelligence amplifier will help doctors to judge patient status more accurately and quickly, thus to get thorough preparation for future treatment (Kononenko 2001). This paper is strongly motivated by the real world problem to predict the potential needs of the patient for pRBC (packed Red Blood Cell) in next few hours from very high resolution vital signs data (ECG and PPG).

We formulate this task as a regular multivariate time series classification problem. Because our data is massive and noisy, the Symbolic Aggregation approXimation (SAX) approach is really a good representation approach for classification (Lin et al. 2003). It has been proved to be effective in several time series data mining tasks such as indexing

(Camera et al. 2010) and visualization (Kumar et al. 2005). Moreover, Bag-of-Pattern makes use of SAX words to encode non-linearity and benefits from invariance to shift (Lin, Khade, and Li 2012). Lin *et al.* reported the previous state-of-the-art results using the One-Nearest-Neighbor classifier (1NN) on UCR Time Series Classification/Clustering databases (Keogh et al.). Oates *et al.* applied SAX and BoP to predict outcomes of traumatic brain injury (Oates et al. 2012a) and explored representation diversity by ensemble voting to further improve classification performance (Oates et al. 2012b).

SAX approach is suitable for univariate time series data. Multivariate time series data is not only characterized by individual attributes, but also by the relationships between the attributes (Bankó and Abonyi 2012). Such information is not captured by the similarity between the individual sequences (Weng and Shen 2008). To deal with the classification problem on multivariate time series, several similarity measurements including Edit distance with Real Penalty (ERP) and Time Warping Edit Distance (TWED) are summarized and tested on several benchmark dataset (Lin J 2012). Recently, a new symbolic representation for multivariate time series classification (SMTS) is proposed. SMTS builds a tree learner with two ensembles to learn the segmentations and a high-dimensional codebook (Baydogan and Runger 2014).

While above methods provide new perspective to handle multivariate data, some are time consuming (e.g. SMTS), some are effective but cannot address the curse of dimensionality (distance on raw data). Can we design such a method to handle the specific type of multivariate *physiological* time series data? We note that strong correlation lies among multivariate time series, especially among the synchronous physiological data. Yu *et al.* made a progress to automatically estimate the reliability of reference heart rates (HRr) derived from ECG and PPG waveforms which is recorded by monitors (Yu et al. 2006). Lu *et al.* compared 5-minute recordings to demonstrate a very high correlation level in the temporal and frequency domains with the nonlinear dynamic analyses between HRV measures derived from PPG and ECG (Lu et al. 2009). They confirmed where HRV measures can be accurately derived in healthy subjects. PPG could also provide the accurate interpulse intervals as a practical alternative to ECG for HRV analysis. Such strong cor-

relations will greatly simplify the feature fusion procedure for us. Some work has been proposed to visualize the correlation among multivariate physiological time series data (Ordóñez, Adviser-Oates, and Adviser-Desjardins 2012).

Another work that our approaches mainly based on is the special pooling structure in Convolutional Neural Networks (CNN), one of the most successful deep learning architectures (LeCun et al. 1989). In this work, we explored to cope with classification problem on synchronous high resolution vital signs data based on the new Pooling SAX algorithms with BoP representations (Pooling SAX-BoP) and Boosting methods (Freund and Schapire 1995). Our contributions are:

- Extending SAX method to multivariate synchronous vital signs time series by applying pooling structures. Relying on the assumption that strong correlation exists among different variables, Pooling SAX can effectively extract the principal information among different synchronous variables.
- Comparing with majority voting (Oates et al. 2012b), we integrated Boosting algorithm with Pooling SAX-BoP representations to further improve the ensemble learning performance on clinical data.
- Proposing a pipeline to process the clinical multivariate vital signs data on predicting patient outcomes.

Pooling SAX-BoP Approaches

The principal idea of SAX is to smooth the input time series using Piecewise Aggregation Approximation (PAA) and assign symbols to the PAA bins. The overall time series trend is extracted as a sequence of symbols.

The algorithm requires three parameters: window length w , number of symbols s and alphabet size a . Different parameters lead to different representations of the time series. Given a normalized time series of length L , we first reduce the dimensionality by dividing it into $\lfloor L/w \rfloor$ non-overlapping sliding windows with skip size 1. Each sliding window is partitioned into s subwindows. Mean values are computed to reduce volume and smooth the noise. Then PAA values are mapped to a probability density function $\mathcal{N}(0, 1)$, which is divided into several equiprobable segments. Letters starting from A to Z are assigned to each PAA values according to their corresponding segments (1).

After discretization and symbolization, Bag-of-Patterns (BoP) are built by a sliding window of length w and convert each subsequence into s SAX words. BoPs catch the features shared in the same structure among different instance and regardless of where they occur. We build our features based on BoP histogram of word counts that is analogous with the bag-of-words (Wang et al. 2013; Lin, Khade, and Li 2012; Baydogan, Runger, and Tuv 2013).

Pooling from Multivariate SAX Word

We proposed a post-processing stage to pool the significant SAX word of each variable with different weighting schemes that is analogous to multiple pooling structures for feature extraction in Convolutional Neural Networks (CNNs). Instead of weights trained by the label, we apply the non-parametric weights to determine the information

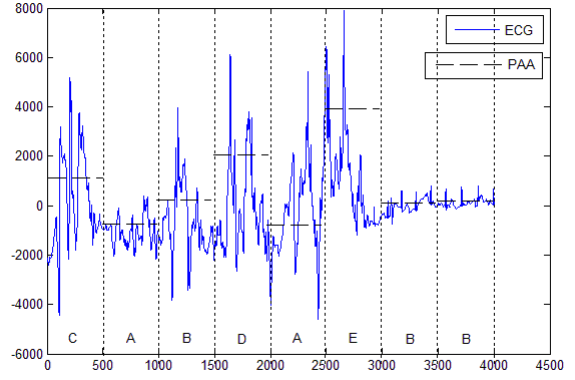


Figure 1: PAA and SAX word for the ECG data. The time series of length 4000 are partitioned into 8 segments. In each segment we compute means to map them to the equiprobable interval. After discretization by PAA and symbolization by SAX, we convert the time series into SAX word sequence CABDAEBB.

density of each variable and pool out the significant words at each time. Finally, we use the pooled-out sequence as single outputs of the multivariate symbolic sequence to build the BoPs. We call such feature extraction and fusion procedure the Pooling SAX-BoP approaches.

Definition. Let \mathcal{X}_k^t denotes a subsequence/bin of time series in the channel k at time t . Operator \mathcal{G} denotes the process of calculating PAA values in each bin, \mathcal{F} is the function to map PAA values to the corresponding SAX word with respect to the standard normal distribution $\mathcal{N}(0, 1)$. \mathcal{W}_k is the non-parametric weights, \mathcal{S} is the pooling output.

Given a subsequence/bin in a sliding window of multiple time series (MTS) $\mathcal{X}_1^t, \mathcal{X}_2^t, \dots, \mathcal{X}_k^t$, the pooling functions of four approaches are given below.

• Max Pooling

$$K_{Max} = \arg \max_k \mathcal{W}_k \cdot \mathcal{F}(\mathcal{G}(\mathcal{X}_k^t))$$

$$\mathcal{S} = \mathcal{F}(\mathcal{G}(\mathcal{X}_{K_{Max}}^t)) \quad (1)$$

Considering the toy example, we extract the SAX word, $[a]$ and $[c]$ from a bin in the bivariate time series. Given $\mathcal{W}_1 = 0.8$, $\mathcal{W}_2 = 0.2$, consider $[a] \times 0.8 = 1 \times 0.8 > 3 \times 0.2 = [c] \times 0.2$, so $\mathcal{S} = [a]$.

• Min Pooling

$$K_{Min} = \arg \min_k \mathcal{W}_k \cdot \mathcal{F}(\mathcal{G}(\mathcal{X}_k^t))$$

$$\mathcal{S} = \mathcal{F}(\mathcal{G}(\mathcal{X}_{K_{Min}}^t)) \quad (2)$$

Considering the same toy example above with the SAX word $[a]$ and $[c]$. Given $\mathcal{W}_1 = 0.8$, $\mathcal{W}_2 = 0.2$, $[c] \times 0.2 < [a] \times 0.8$, then $\mathcal{S} = [c]$.

• Max-Min Pooling

After extracting two significant SAX words (their variable index K_{Min} and K_{Max}) from pooling function in Equation. (1) and (2),

$$\mathcal{S} = [\mathcal{F}(\mathcal{G}(\mathcal{X}_{K_{Max}}^t)), \mathcal{F}(\mathcal{G}(\mathcal{X}_{K_{Min}}^t))] \quad (3)$$

Table 1: CV error rates on two standard datasets

	ECG	wafer
Pooling SAX Max	0.115	0.0293
Pooling SAX Min	0.115	0.0168
Pooling SAX Max-Min	0.115	0.0242
Pooling SAX 3M	0.115	0.0242
SMTS	0.134	0.01
Euclidean	0.1778	0.0833
DTW(full)	0.1889	0.0909
DTW(window)	0.1722	0.0656
EDR	0.2	0.3131
ERP	0.1944	0.0556
LCSS	0.1278	0.1363
LCSS Relaxed	0.1278	0.1091
TWED	0.1278	0.0318

Considering the same example with the maximum pooling word $[a]$ and the minimum pooling word $[c]$. then $S = [a, c]$.

- **Max-Min-Mean Pooling (3M Pooling)**

Still this need to figure out the significant SAX word index K_{Min} and K_{Max} from pooling function in Equation. (1) and (2), we consider the weighted average of the SAX word among different channels as

$$\begin{aligned} S_{Mean} &= \left[\sum_{k=1}^K \frac{\mathcal{W}_k}{Z} \cdot \mathcal{F}(\mathcal{G}(\mathcal{X}_k^t)) \right] \\ Z &= \sum_{k=1}^K \mathcal{W}_k \end{aligned} \quad (4)$$

Then

$$S = [\mathcal{F}(\mathcal{G}(\mathcal{X}_{K_{Max}}^t)), S_{Mean}, \mathcal{F}(\mathcal{G}(\mathcal{X}_{K_{Min}}^t))] \quad (5)$$

Given two synchronous SAX words $[a]$ and $[c]$ with the weights $\mathcal{W}_1 = 0.8$, $\mathcal{W}_2 = 0.2$, from Equation. (4), $S_{Mean} = [1 \times 0.8 + 3 \times 0.2] = 2 = b$, then $S = [a, b, c]$

Above four pooling approaches are actually inspired from the pooling architecture in CNNs. Max pooling in CNN attempts to extract the significant weight vectors with respect to the labels to achieve translation invariance. For multivariate time series, we suppose to pool out the most significant channels with more information density. Max/Min Pooling provide us the translation invariance cross multiple channels at the same temporal interval. Max-Min pooling and 3M pooling are much like the multiple K-pooling. The motivation stems from the significance of the extreme values in time series. 3M pooling combines the average pooling with multiple K-pooling. In 3M pooling, we suppose to observe how the weighted average value regulates the behavior of the BoP representations together with max and min values.

The weight scheme \mathcal{W}_k is a series of non-parametric weights. They decide the significance level of each SAX word in different channels at the same temporal interval. We have multiple choices to define \mathcal{W}_k . Entropy (Shannon 2001) concerns more about the information density of

Table 2: Test error rates on two standard dataset

	ECG	wafer
SMTS	0.182	0.035
MTSBF	0.165	0.015
Pooling SAX Max	0.16	0.02
Pooling SAX Min	0.18	0.033
Pooling SAX Max-Min	0.20	0.031
Pooling SAX 3M	0.18	0.039

each channel, average KL-divergence (Kullback and Leibler 1951) measures the difference between the object variable and other variables. Permutation Entropy (Bandt and Pompe 2002) evaluates the complexity of a given time series. All these measurements are nonparametric but have different computation complexity and focus on different aspects. In this paper, we mainly explore the physiological time series data. As aforementioned discussion, strong correlations are always observed among different channels/variables. We care more about the information density in each synchronous variables as they tend to have more significant regulations in synchronous data. Our weight scheme is defined as the rescaled variance:

$$\mathcal{W}_k = \frac{\sum_{t=1}^{L_k} (\mathcal{X}_k^t - \bar{\mathcal{X}}_k)^2}{L} \quad (6)$$

In the above equation, \mathcal{X}_k^t is rescaled into the interval $[0, 1]$. Rescaled variance evaluates the information density regardless of the magnitude in each channel. This weight scheme regulates the pooling behavior to extract the significant features while preferring the channels with more information.

After pooling out the single sequence of SAX word from the multivariate time series, we build the Bag-of-Patterns to classify the multivariate time series with 1NN classifier.

Experiments on Standard Dataset and High Resolution Vital Signs Data

Lin *et al.* compared multiple distance measurements for classification on multivariate time series and use 10 fold cross-validation (CV) to evaluate the performance (Lin J 2012). They chose four multivariate time series datasets. Two of them are too short to use SAX approaches (the average length of "AUSLAN" is 57 and the lengths of "Japanese Vowels" range from 7 to 29). We use the other two dataset and compare the CV error rate with other approaches (include the current state-of-the-art approach SMTS (Baydogan and Runger 2014)) in Table. 1. Baydogan *et al.* split these two dataset as training and testing. They reported the performance on the training-testing manner with their SMTS and MTSBF methods. We also train our approach with CV on the training set and compare the performance on the test set (Table. 2). Our methods are proved to be quite competitive with the SMTS approach while our approaches are simpler (average running time is 3 hours) without any ensemble framework. However, SMTS needs two ensembles to learn a tree-based codebook. Their average running time is reported to be 18 hours.

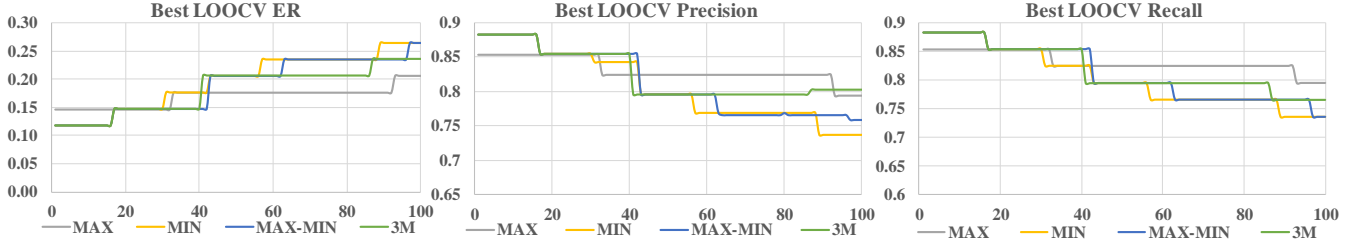


Figure 2: Ranked classification error rate/precision/recall of pooling SAX-BoP on 100 bootstrap vital signs dataset.

In the next experiment, we test our approach on the bivariate high resolution vital signs data. The clinical data is collected in the University of Maryland, Medical School. All the data is anonymous in order to protect the privacy. The ECG and PPG data of 556 patients were recorded in 68 to 128 minutes long with 240 Hz sampling rate. Default value (e.g. -31556 when there is no input) and missing value are allowed. The length of 237 samples are less than 128 minutes. The data is quite massive with more than 1.7 million data points. The labels indicate if a patient needs blood transfusion of pRBC or not in the next 6 hours. The data is highly skewed with only 17 positive samples.

Considering the vital signs data is highly periodical, we preprocess the data second by second. Among 7680 seconds, we get rid of any interval with the default or missing value. This may lead to information loss, because some seconds only contain few missing number or default value. However, inconsistency caused by missing number and default value may exert deeper hazard due to the noise and false information. Based on the fact that the normalized time series has approximate Gaussian distribution (Lin et al. 2003), we applied variance filter to further regulate the outliers. According to the 3-Sigma rule (Pukelsheim 1994), if a value locates beyond three times deviations away from the mean, the probability of that point incurring is naturally lower than 0.27%. Thus, we truncate these outliers to the lower and upper bounds. What we need is the overall trend encoded in the time series, the last step of preprocessing is calculating means in each second to reduce the volume and keeping the overall trend analogously to PAA.

To guarantee the results on our dataset is not biased, we build 100 new balanced datasets using bootstrap (Manly 2006) by keeping all the 17 positive samples fixed and randomly choosing 17 negative samples with replacement. To compare the performance of the Pooling SAX-BoP approaches on clinical data, we also test standard SAX-BoP approach on each single variable of the vital signs data and report the statistics of the best 1NN classifier error rate, precision and recall with Leave-One-Out cross validation (LOOCV) on the 100 bootstrap dataset (Table . 3).

Figure. 2 shows the ranked curve of classification error rate, precision and recall of four pooling SAX-BoP approaches on 100 resampled bootstrap datasets. In our experiments, except for the Max Pooling, all other three approaches has the equivalent best error rate (0.117) among all

Table 3: The LOOCV statistics of the best performance for standard SAX-BoP and multivariate Pooling SAX-BoP on 100 bootstrap dataset

	Error Rate	Precision	Recall
PPG	$0.192 \pm (0.037)$	$0.813 \pm (0.043)$	$0.808 \pm (0.037)$
ECG	$0.256 \pm (0.038)$	$0.748 \pm (0.037)$	$0.744 \pm (0.038)$
MAX	$0.169 \pm (0.017)$	$0.831 \pm (0.017)$	$0.831 \pm (0.02)$
MIN	$0.196 \pm (0.049)$	$0.807 \pm (0.048)$	$0.804 \pm (0.047)$
MAX-MIN	$0.188 \pm (0.048)$	$0.813 \pm (0.047)$	$0.811 \pm (0.046)$
3M	$0.181 \pm (0.040)$	$0.824 \pm (0.040)$	$0.818 \pm (0.034)$

dataset. Although Max Pooling cannot reach to the best performance as others, it is more stable with slight oscillation in the performance curve. It also demonstrates better average statistics with small standard deviation. Our experiments on the clinical physiological data imply that Pooling SAX-BoP approaches improve the expressive power of the BoP representations and enhance the classification performance. They not only demonstrate the competitive state-of-the-art performance on standard datasets, but also work well to solve the real world problems.

Ensemble Learning on BoP Representations Using Boosting Algorithm

SAX reduces the dimensionality but also drops much details in the raw data. The Pooling SAX methods attempt to extract the significant SAX words while preserving the core information. This means SAX based approaches run the risk of information loss where the key structures in specific time series might be discarded, thus leads to the samples to be misclassified. Some work has been proposed to exploit representational diversity for time series classification via ensembles of the representations (Oates et al. 2012b). We randomly select a balanced dataset to test majority voting approach as a baseline. That is, a subset of those classifiers is chosen and allowed to vote on the class labels for test set instances, with the unweighted majority used to predict the class label. We chose the subset by keeping only those representations with top K lowest error rates of the best representation as measured by 1NN error on the training set.

In Figure. 3 (a), we observe the effect of information loss in the voting results. As the number of voting agents increase from 10 to 70, there is no change in LOOCV error rates for Max Pooling SAX. The enhancement of performance on other pooling structures is also not so clear. That is, informa-

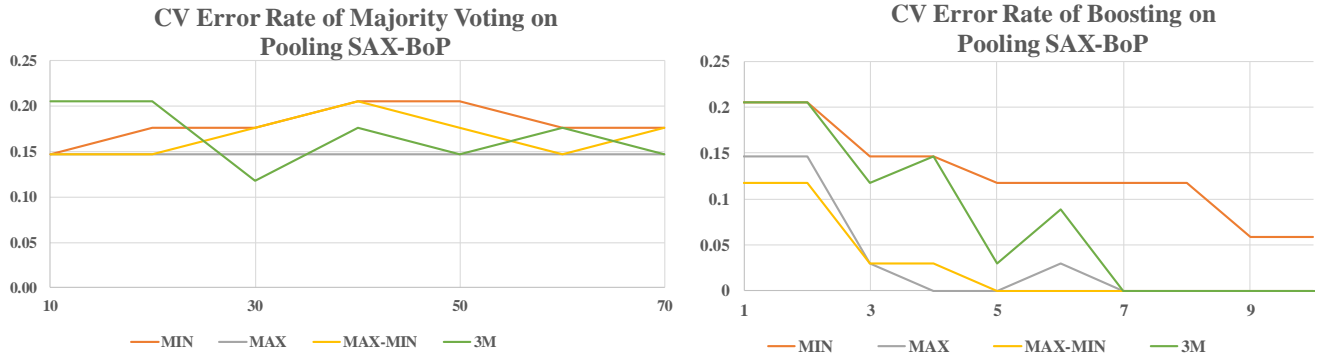


Figure 3: LOOCV Error Rate of a) majority voting (left) and b) Boosting (right) with Pooling SAX-BoP approaches on vital signs data.

tion loss leads to the failure to interpret key features among the top Pooling SAX-BoP representations.

Boosting on Pooling SAX-BoP Representations

To address misclassification caused by feature missing in the Pooling SAX-BoP approaches, we apply Boosting algorithm to build a non-linear classifier (Freund, Schapire, and others 1996). Boosting adaptively changes the sample weights according to previous classification results to focus on the toughest samples. The missing feature dimension caused by information loss is hit by the larger weight during the iterative process. Instead of increasing the dimensionality of the feature space in kernel methods, we use Boosting to tune the linear classification hyperplane of several weak classifiers into a nonlinear classification hyperplane by weighted summation. Despite of missing dimension, the nonlinear hyperplane potentially classify some tough samples in the linear situations.

We apply a slightly modified version of Boosting algorithm for SAX-BoP representations to classify the balanced vital signs dataset. The trick is to combine each SAX-BoP patterns with a 1NN classifier as a weak learner in Boosting algorithm. What Boosting does for 1NN classifier is to create an ensemble of models with locally modified distance weighting (Athitsos and Sclaroff 2005). After approximately 10 turns voting, the converged performance is significantly enhanced (Figure 3 (b)).

Recall that we preprocessed high resolution vital signs data by averaging in each second interval. If voting through the diversity of BoP representation enrich the information and enhance the performance, multiple preprocessing frequencies can also capture different temporal information in the vital signs data, respectively. In preprocessing stage, we calculate the mean value in each 1, 0.5 and 0.3 second interval and combine the Pooling SAX-BoPs on these three preprocessing frequencies together into one large dataset. Because better weak learner will be selected in each iteration from the samples of three resampling frequencies, different resampling rate will mix various temporal diversity into Boosting algorithm to further enhance the classification performance and accelerate the converge speed (Figure.4).

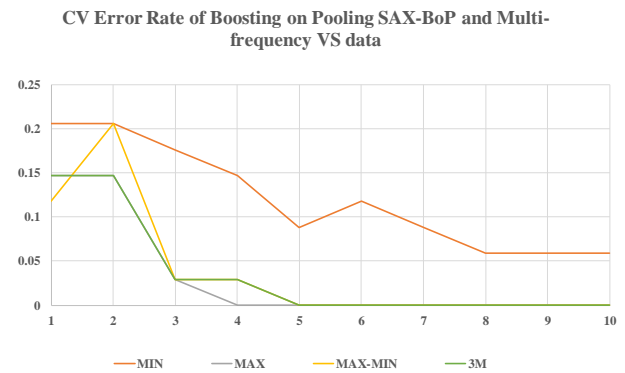


Figure 4: LOOCV Error Rate of Boosting algorithm with Pooling SAX-BoP approaches on the vital signs data of multiple frequency.

Boosting runs the risk of overfitting. The direct application of VC theory shows that boosting can work well if providing simple weak classifiers which satisfy the weak learning condition, and if running for enough but not too many rounds (Schapire 2013). We update the error rate from the optimal SAX-BoP representation and 1NN classifiers with the higher accuracy than 50%. The convergence curve in Figure. 3 (b) shows that Boosting algorithm in our case does not need many iterations (About 10 turns) to converge to its stable performance. We applied the fully trained Boosting classifier on the balanced dataset of multiple frequency to classify the fully 556 samples. The Boosting classifier achieved 4.856% error rate. Meanwhile, predicting using the trained Boosting classifier is very fast and convenient for real-time classification of physiological data in a second-by-second manner.

Conclusion and Future Work

In this paper, we proposed the pooling SAX-BoP with Boosting algorithm to solve classification problem on the multivariate vital signs time series data. The experiments on two standard datasets and the clinical vital signs data

demonstrate the effectiveness and efficiency of the Pooling SAX-BoP approach. Our approach achieved competitive results with two current state-of-the-art multivariate time series classifiers. Instead of majority voting, Boosting algorithm is applied to significantly improve the performance.

Future work includes exploring the correlation between multivariate vital signs data through statistics and feature selection technique to further improve the performance. We also plan to explore more clinical vital signs data to predict patient outcomes.

References

- Athitsos, V., and Sclaroff, S. 2005. Boosting nearest neighbor classifiers for multiclass recognition. In *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, 45–45. IEEE.
- Bandt, C., and Pompe, B. 2002. Permutation entropy: a natural complexity measure for time series. *Physical Review Letters* 88(17):174102.
- Bankó, Z., and Abonyi, J. 2012. Correlation based dynamic time warping of multivariate time series. *Expert Systems with Applications* 39(17):12814–12823.
- Baydogan, M. G., and Runger, G. 2014. Learning a symbolic representation for multivariate time series classification. *Data Mining and Knowledge Discovery* 1–23.
- Baydogan, M. G.; Runger, G.; and Tuv, E. 2013. A bag-of-features framework to classify time series. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 35(11):2796–2802.
- Camerra, A.; Palpanas, T.; Shieh, J.; and Keogh, E. 2010. isax 2.0: Indexing and mining one billion time series. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, 58–67. IEEE.
- Freund, Y., and Schapire, R. E. 1995. A decision-theoretic generalization of on-line learning and an application to boosting. In *Computational learning theory*, 23–37. Springer.
- Freund, Y.; Schapire, R. E.; et al. 1996. Experiments with a new boosting algorithm. In *ICML*, volume 96, 148–156.
- Keogh, E.; Zhu, Q.; Hu, B.; Hao, Y.; Xi, X.; Wei, L.; and Ratanamahatana, C. The ucr time series classification/clustering, 2011. *Homepage*: http://www.cs.ucr.edu/~eamonn/time_series_data.
- Kononenko, I. 2001. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in medicine* 23(1):89–109.
- Kullback, S., and Leibler, R. A. 1951. On information and sufficiency. *The annals of mathematical statistics* 79–86.
- Kumar, N.; Lolla, V. N.; Keogh, E. J.; Lonardi, S.; and Ratanamahatana, C. A. 2005. Time-series bitmaps: a practical visualization tool for working with large time series databases. In *SDM*, 531–535. SIAM.
- LeCun, Y.; Boser, B.; Denker, J. S.; Henderson, D.; Howard, R. E.; Hubbard, W.; and Jackel, L. D. 1989. Backpropagation applied to handwritten zip code recognition. *Neural computation* 1(4):541–551.
- Lin, J.; Keogh, E.; Lonardi, S.; and Chiu, B. 2003. 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*, 2–11. ACM.
- Lin J, Williamson S, B. K. D. D. 2012. *Pattern recognition in time series*. Chapman & Hall, To appear.
- Lin, J.; Khade, R.; and Li, Y. 2012. Rotation-invariant similarity in time series using bag-of-patterns representation. *Journal of Intelligent Information Systems* 39(2):287–315.
- Lu, G.; Yang, F.; Taylor, J.; and Stein, J. 2009. A comparison of photoplethysmography and ecg recording to analyse heart rate variability in healthy subjects. *Journal of medical engineering & technology* 33(8):634–641.
- Manly, B. F. 2006. *Randomization, bootstrap and Monte Carlo methods in biology*, volume 70. CRC Press.
- Oates, T.; Mackenzie, C. F.; Stansbury, L. G.; Aarabi, B.; Stein, D. M.; and Hu, P. F. 2012a. Predicting patient outcomes from a few hours of high resolution vital signs data. In *Machine Learning and Applications (ICMLA), 2012 11th International Conference on*, volume 2, 192–197. IEEE.
- Oates, T.; Mackenzie, C. F.; Stein, D. M.; Stansbury, L. G.; Dubose, J.; Aarabi, B.; and Hu, P. F. 2012b. Exploiting representational diversity for time series classification. In *Machine Learning and Applications (ICMLA), 2012 11th International Conference on*, volume 2, 538–544. IEEE.
- Ordonez, P.; Adviser-Oates, T.; and Adviser-Desjardins, M. 2012. *Multivariate time-series analysis of physiological and clinical data*. University of Maryland at Baltimore County.
- Pukelsheim, F. 1994. The three sigma rule. *The American Statistician* 48(2):88–91.
- Schapire, R. 2013. Explaining adaboost. In Scholkopf, B.; Luo, Z.; and Vovk, V., eds., *Empirical Inference*. Springer Berlin Heidelberg. 37–52.
- Shannon, C. E. 2001. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review* 5(1):3–55.
- Wang, J.; Liu, P.; She, M. F.; Nahavandi, S.; and Kouzani, A. 2013. Bag-of-words representation for biomedical time series classification. *Biomedical Signal Processing and Control* 8(6):634–644.
- Weng, X., and Shen, J. 2008. Classification of multivariate time series using locality preserving projections. *Knowledge-Based Systems* 21(7):581–587.
- Yu, C.; Liu, Z.; McKenna, T.; Reisner, A. T.; and Reifman, J. 2006. A method for automatic identification of reliable heart rates calculated from ecg and ppg waveforms. *Journal of the American Medical Informatics Association* 13(3):309–320.