

Classification of patients using novel multivariate time series representations of physiological data

Patricia Ordóñez*, Tom Armstrong†, Tim Oates*, and Jim Fackler‡

*Computer Science and Electrical Engineering Department
University of Maryland, Baltimore County, Baltimore, MD

Email: patti.ordonez@umbc.edu, oates@cs.umbc.edu

†Mathematics and Computer Science Department
Wheaton College, Norton, MA

Email: tarmstro@wheatoncollege.edu

‡Pediatric Anesthesiology and Critical Care Medicine
Johns Hopkins University School of Medicine, Baltimore, MD
Email: fackler@jhmi.edu

Abstract—In this paper we present two novel multivariate time series representations to classify physiological data of different lengths. The representations may be applied to any group of multivariate time series data that examine the state or health of an entity. *Multivariate Bag-of-Patterns* and *Stacked Bags-of-Patterns* improve on their univariate counterpart, inspired by the *bag-of-words* model, by using multiple time series and analyzing the data in a multivariate fashion. We also borrow techniques from the natural language processing domain such as term frequency and inverse document frequency to improve classification accuracy. We introduce a technique named *inverse frequency* and present experimental results on classifying patients who have experienced acute episodes of hypotension.

I. INTRODUCTION

The sophistication of instrumentation to collect high frequency vital signs data and the resulting volume of data are greater now than at any point in the history of medicine. Given this fact, the information overload that medical providers experience may inhibit the diagnostic process [1]. Providers are expected to examine large amounts of data and identify correlations between dozens, if not hundreds, of variables based on, in part, their own clinical experience to detect significant medical events. Data with low dimensionality (e.g., laboratory tests, prior vital signs monitoring systems) presented fewer analysis challenges. For example, a recent study found that humans successfully process up to four independent variables accurately and efficiently when data are presented in bar graphs. However, when a fifth independent variable is introduced, performance drops to that of chance [2].

Visualizations of composite variables and univariate data may aid providers. The state-of-the-art in data visualization for vital signs assists the provider in analyzing the information through a table or plot of values for a particular parameter over time (i.e., a time series). Recently, these displays have been augmented with MTSA (Multivariate Time Series Amalgams) Visualizations based on star plots [3]. We are interested in making available additional information from large databases of patient vital signs to assist the development of clinical decision support systems in visual analytics.

Current techniques focus on making more user-friendly interfaces to build complex queries [4], [5], [6]. Furthermore, these queries are often based on traditional baselines and thresholds for variables, which indicate “normal” ranges of values in a healthy patient. These values are generally based on age, gender, or weight, and were established (often decades ago) by measuring these variables in a large population of patients. While these values are appropriate for many healthy people, they are not for others because of differences in lifestyle, diseases, genetic makeup, and environment.

With the development of electronic medical records, it is now possible to use a patient’s medical history to automate the personalization of baselines and thresholds while incorporating vital signs data from similar patients. Automated techniques for discovering the correlations between variables and other patients would assist the provider in making a diagnosis and help in identifying hidden patterns within the data.

We have taken the first step in classifying a patient’s physiological data by creating two representations of multivariate time series that compare physiological data of varying lengths by structure instead of by shape. In other words, the representations and techniques in this paper classify multivariate vital signs data by breaking the time series into small temporal structures and determining similarity based on the frequency of these structures as opposed to the shape of the time series or the ranges of different variables.

In the following sections, we review background on time series discretization methods and approaches to classification of univariate data (including single variable vital signs data). Next, we discuss our representations, Multivariate Bag-of-Patterns and Stacked Bags-of-Patterns, for classifying multivariate time series data. Then, we present some of the natural language processing techniques we incorporated in the classification of the data including one unconventional technique inspired by the work of Saeed and Mark [7]. Finally, we conclude with empirical results of our Multivariate Bag- and Stacked Bags-of-Patterns representation on multivariate data versus univariate data on publicly available multivariate

physiological data and point toward future work.

II. BACKGROUND

Efficient processing of multivariate time series databases for querying [8], [9], segmenting [10], and discovering patterns or anomalies [11] are challenging problems [12]. In most scenarios, given sufficient space and time, an exhaustive account of the entire time series database lends itself to solving many of these problems. However, in many time series domains, including the medical domain, the actual space and time requirements preclude delayed or real-time processing. To address this challenge, several univariate approaches employ randomized algorithms [13] or uniform scaling in the time dimension [14]. Work on mining multivariate time series data often suffers from the curse of dimensionality. One approach to mitigating this challenge extends the univariate randomized algorithm to approximate which dimensions to consider [15].

Finding the similarity between two time series is a non-trivial task simply because of the temporal characteristics of the data. Two time series may be identical in shape, but not be synchronized in time. If comparing point to point raw values using the standard metric of Euclidean distance, such time series would be considered very different. For this reason, Dynamic Time Warping (DTW), a dynamic programming scheme for comparing two time series by locating the two closest points within a specified window is more commonly used to compare time series by shape. Other methods focus on translating the times into a mathematical expression and compare the coefficients used to create the expression such as wavelet [16] and Fourier transformations. Yet other methods ignore the shape and focus on the structural similarity instead. Our work is based on the latter type of method and we focus on work related to this approach.

To simplify data mining for time series data, we considered using several alternative time series representations, including Discrete Fourier Transformation [17], Discrete Wavelet Transformation [18], and Piecewise Linear Approximation [12]. We chose to use SAX [19] because of its ease of computation and comprehension, its ability to reduce the data's dimensionality, and it is unique in guaranteeing a lower bound on Euclidean distances between time series. The latter characteristic allows the use of data mining algorithms on a symbolic representation that then give similar results to the original time series.

A. Symbolic Aggregate approXimation

The Symbolic Aggregate approXimation (SAX), and more recently iSAX [20], has gained traction as the *de facto* representation for time series in the large space of representation choices [21]. The usefulness of SAX in detecting motifs and anomalies in univariate time series data is particularly appealing in medicine where anomalies in physiological data often pertain to medically significant events.

The algorithm to convert a real valued time series into a SAX string representation has four parameters: window length (w), number of symbols per window (s), alphabet size (a), and numerosity reduction (nr). Prior to discretizing the univariate

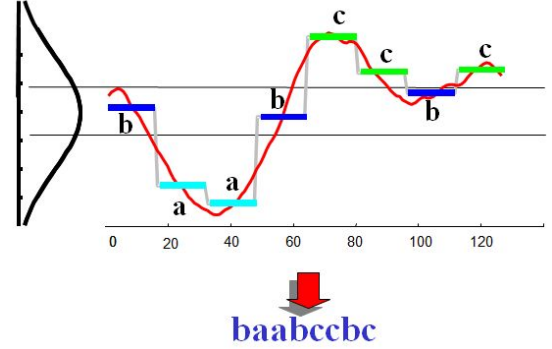


Fig. 1. Normalized time series converted to PAA and annotated with SAX symbols ($a = 3$; $s = 8$; $w = 128$), [19]

time series using SAX, the data are normalized such that the average is 0 and the standard deviation is 1.

SAX then divides the time series (shown as a curved line in Figure 1) of length w into s segments of equal width along the x-axis. The time series in Figure 1 has $w = 128$ and $s = 8$. The average of the values within each segment is used as the value for the segment in a new discretized time series. This representation of time series is known as Piecewise Aggregate Approximation (PAA, represented by the colored horizontal lines in Figure 1). SAX then examines the Gaussian distribution of values and divides the distribution into equal parts, depending on the desired alphabet size. For example, for an alphabet size of 3, the separations occur at $-.43$ and $.43$ in a normalized Gaussian distribution, as shown by the straight horizontal gray lines in Figure 1. The PAA value for each segment is then converted into a symbol, depending on the section of the distribution into which the PAA value falls (labels a, b and c in Figure 1). SAX then slides this window across the time series to create the representation depending on the value of the numerosity reduction parameter, nr , which is described in the next section.

B. Finding Structural Similarity Univariately

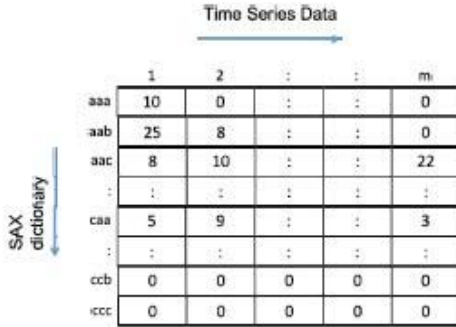
Much of the previous research on time series analysis has focused on univariate time series data [12]. Our interest is in examining multiple time series per individual where the time series may be of different lengths and the ranges for each variable may be orders of magnitude different. To address this, we consider an approach that focuses on comparing the structural similarity of univariate time series, Bag-of-Patterns (BoP) [22], and how we might extend it to a multivariate setting. BoP outperforms existing techniques for classification, clustering and anomaly detection of univariate time series. The BoP representation takes the SAX string of the time series in small local areas throughout the time series, which can be equated to the words in a document, and counts the occurrence of these words to create a vector of word frequencies. By doing so, it paves the way for the comparison of two time series in the same way that you would compare two documents in a

bag-of-words style, by examining the histograms of their word frequencies.

In BoP, each time series is converted into a frequency vector of length a^s . The vectors are calculated by having a window of length w slide over the entire time series and converting the subsequences into a SAX representation, a word. The frequency of the SAX words is stored in the vector. Identical consecutive words are only counted once when numerosity reduction is set, to avoid matching based on recurring patterns (like stop words in a document).

The latest SAX implementation actually has four different levels of numerosity reduction: from no numerosity reduction (all words are counted), to default (described above), to advanced where the word is recorded only if the *MinDist* [19] between the words is greater than zero, to maximum numerosity reduction where the subsequence is recorded only if it is not monotonic.

Figure 2 displays the frequency vectors for a collection of m time series. Distance between the vectors can be calculated using any conventional distance metric. An example of using BoP in hierarchical clustering is shown in Figure 3 [22]. Each time series is followed by a histogram of its frequency vector. Euclidean distance was used as the similarity metric. The resulting dendrogram is in Figure 3.



	1	2	...	m
aaa	10	0	...	0
aab	25	8	...	0
aac	8	10	...	22
caa	5	9	...	3
ccb	0	0	...	0
ccc	0	0	...	0

Fig. 2. BoP representations of m time series [22]

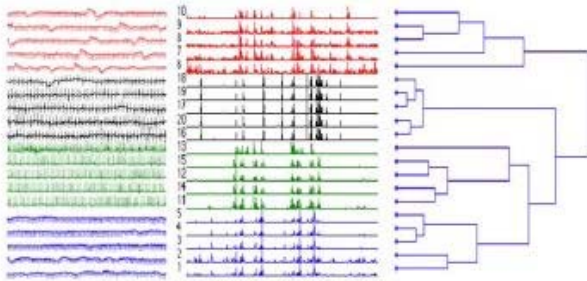


Fig. 3. The histogram of the BoP representation of 20 time series and the resulting dendrogram after hierarchical clustering [22]

BoP for classification of univariate time series works as follows: given an alphabet size, a , and the number of symbols

in a subsequence, s , the length of the frequency vector is equivalent to the number of “SAX words” in the vocabulary, a^s . Because many words are not found in the time series, a data structure with compressed storage is used. In Figure 2, the BoP representations for m time series are displayed in a matrix. Because all of the time series are of the same length, numerosity reduction takes care of weighting the most unusual subsequences more and the number of times that a word is displayed in the SAX representation is sufficient for comparing the time series.

III. FINDING STRUCTURAL SIMILARITY MULTIVARIATELY

We expand on the BoP representation by tailoring it to multivariate medical data, where emphasizing unusual patterns is critical to classifying the patients correctly. For this purpose, we use techniques from natural language processing [23] and apply them to the multivariate BoP representations.

A. Multivariate Bag-of-Patterns

Multivariate Bag-of-Patterns (Multivariate BoP) is a representation for multivariate time series data that is designed to capture the relationships between the time series over time. It was inspired by the idea that the complexity of the human organism is better captured by analyzing how organs relate to one another than by analyzing the organs individually, and thus is tailored for multivariate time series that represent the state or health of a complex entity.

Multivariate BoP takes multiple univariate times series and converts them into their SAX representation individually. There is no numerosity reduction so that the SAX representations for the different parameters are the same length for one patient. Thus, the trends across the parameters are captured in time by creating multivariate words, “MV words”, that represent the state of the patient in one time interval. That is, each “MV word” contains the SAX tokens from the univariate time series at a particular time. Once the “MV words” are created, we apply the Bag-of-Patterns approach and create a frequency vector for the patient. See Figure 4 for a visualization of processing a single patient’s data with 2 variables.

B. Stacked Bags-of-Patterns

The second of our multivariate time series representations is *Stacked Bags-of-Patterns* (Stacked BoP). As in Multivariate BoP, each univariate time series in the patient is converted to its SAX representation. Unlike Multivariate BoP, each is treated as an individual BoP instance and all the BoP vectors are concatenated together into a single, Stacked BoP vector. Thus, the vector contains the frequencies of “SAX words” over all the parameters. See Figure 5 for a visualization of processing a single instance with n variables.

C. Adapted Natural Language Processing Techniques

The univariate BoP approach has proven effective with time series of the same length. However, we suspect that when tested with time series of different lengths, as is the case with

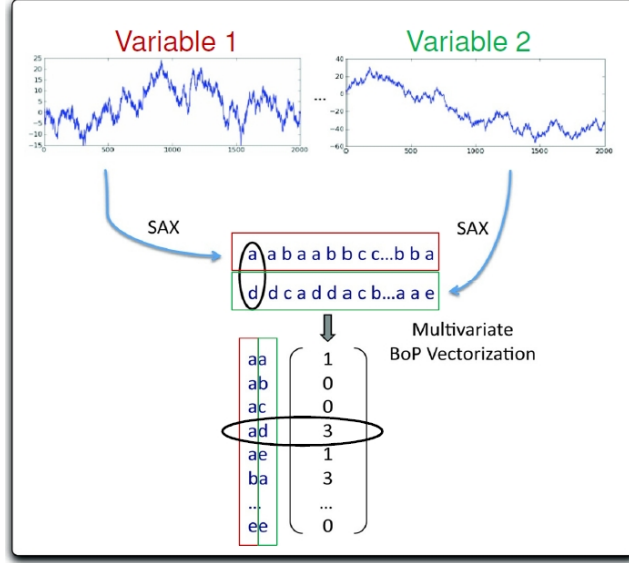


Fig. 4. Multivariate Bag-of-Patterns for a single 2-variable instance

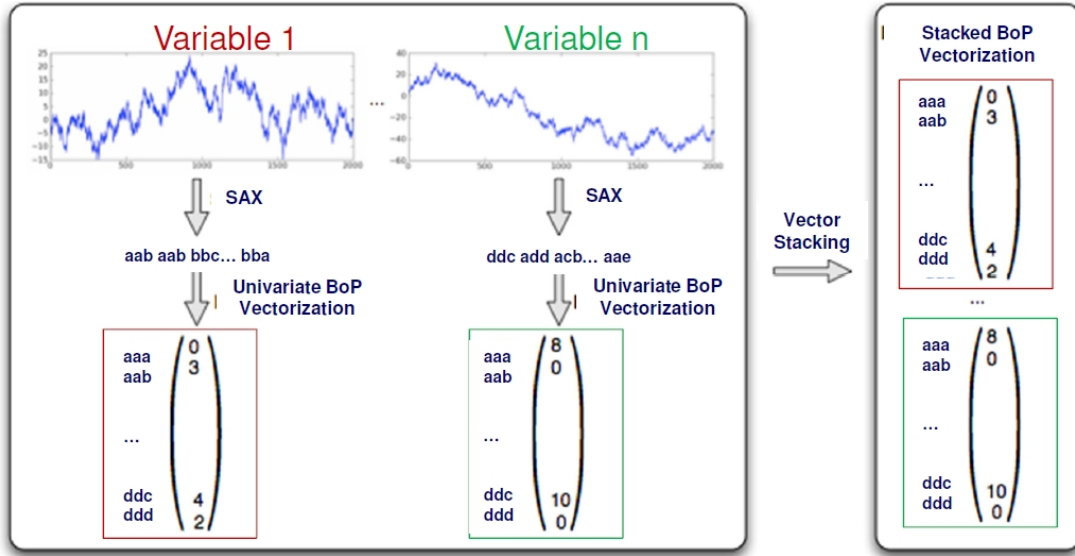


Fig. 5. Stacked Bags-of-Patterns for a single n -variable instance

most patient vital signs data, BoP may prove less effective. To resolve this, we have incorporated the following language processing techniques: Term Frequency (TF; Eqn 1), Inverse Document Frequency (IDF; Eqn 2), and Inverse Frequency (IF; Eqn 3). A combination of the last two techniques was implemented in IDF-IF (Eqn 4).

In equations 1, 2, 3, and 4, x_i is an “MV word” or “SAX word”, m is the number of words in the vocabulary ($a^s * p$ for Stacked BoP and a^p for Multivariate BoP where p is the number of parameters), and n is the number of instances (i.e., number of patients). $f_n(x_i)$ is the number of instances that contain word x_i . $f_a(x_i)$ is the number of occurrences of word

x_i in all instances. Each of the multivariate BoP vectors for a single instance, $patient_j$, is weighted term-by-term using combinations of the following:

$$TF(x_i) = \frac{f_{patient_j}(x_i)}{\sum_{i=1}^m f_{patient_j}(x_i)} \quad (1)$$

$$IDF(x_i) = \log\left(\frac{n}{f_n(x_i)}\right) \quad (2)$$

$$IF(x_i) = \log\left(\frac{n}{f_a(x_i)}\right) \quad (3)$$

$$\text{IDF} - \text{IF}(x_i) = \log\left(\frac{n}{f_n(x_i) * f_a(x_i)}\right) \quad (4)$$

Bag-of-Patterns (BoP) is based on the Vector Space Model (VSM) [24] (a.k.a. “bag of words” model) for comparing similar documents. In this model, two documents are compared based on the frequency of the common words within them. However, words that are very common such as articles and conjunctions that do not have to do with the content of the documents and thus do not contribute to their similarity are omitted. The frequency of a word between two documents is not a good measure of similarity unless more weight is given to the unusual words they share. Inverse Document Frequency gives more weight to unusual words so that similarity between two documents is influenced by the number of uncommon words they share as displayed in Eqn 5.

$$\text{VSM}(\text{patient}_j) = [\text{TF}(x_1)\text{IDF}(x_1), \dots, \text{TF}(x_m)\text{IDF}(x_m)] \quad (5)$$

Whereas IDF typically describes how infrequently a word is found within a corpus of documents, Inverse Frequency describes how infrequently a pattern is found in a collection of documents. Thus, the more unusual the pattern in all documents, the more weight the pattern is given. Both IF and IDF-IF were inspired by the work of Saeed and Mark ([7]). The multivariate BoP approach allows us to perform patient “word” comparisons like those performed on documents. The idea in doing so is to capture similarity between patients by emphasizing the unusual patterns they share in the hope of finding patients sharing the same illness and/or disease since these are not the normal states of the population of patients.

IV. EXPERIMENTS

The goal of our experiments was to compare the univariate BoP representations to the multivariate ones. We decided to use one nearest neighbor classification (1-NN) because it is one of the most common and easily understood algorithms, and it was used in Lin’s Bag-of-Patterns paper [22]. Thus, the results in the paper were calculated using 1-NN classification with Euclidean distance and leave-one-out cross validation.

A. Stacked Bags-of-Patterns

To evaluate our Stacked BoP method, we constructed a synthetic multivariate dataset derived from a real univariate dataset. The basis for the dataset is one of the evaluation datasets used to demonstrate the univariate BoP approach [22]. We artificially constructed a bivariate version of the data by selecting pairs of univariate instances. The original 26 univariate instances are of length 1000 and come from a variety of time series domains contained in the UCR Time Series Archive [25]. There are a total of 13 classes – two univariate instances per class – in the original univariate data set.

In our synthetic data set, we built 13 classes of three bivariate time series by creating all 2-combinations of each univariate class. For example, if class 1 contained instances 1a and 1b, our bivariate class would contain three bivariate

instances (1a1b-1a1a-1b1b). Then we created 78 (sum of consecutive integers between 1 and 12) bivariate classes of four instances by creating all 2-combinations of two univariate classes. For example, combining univariate classes 1 and 2 would create a new class of four bivariate time series (1a2a-1a2b-1b2a-1b2b). In the end, we built a dataset containing 351 (4*78 + 3*13) bivariate instances in 91 (13 + 78) classes. To convert the raw time series data into the SAX string representation, we used a window size of 100, number of subsequence symbols of 6, SAX alphabet of size 4, and different levels of numerosity reduction.

The results of 1-NN classification on the raw time series (baseline), Stacked BoP, Stacked BoP with Term Frequency, Stacked BoP with Term Frequency-Inverse Document Frequency, and Stacked BoP with Term Frequency-Inverse Document Frequency-Inverse Frequency are reported in Table I. All versions of Stacked BoP outperform the raw time series data baseline.

B. Experiments on real data sets

To evaluate our approaches on real world vital signs data, we used two of the challenge data sets available from PhysioNet [26]. We selected Test Set B and the Training Set of the “PhysioNet/Computers in Cardiology Challenge 2009: Predicting Acute Hypotensive Episodes”. An acute hypotensive episode (AHE) is an episode of low blood pressure that PhysioNet describes as:

“Among the most critical events that occur in intensive care units (ICUs), acute hypotensive episodes require effective, prompt intervention. Left untreated, such episodes may result in irreversible organ damage and death. Timely and appropriate interventions can reduce these risks.” [27].

We began evaluating the representations using the clinical data for Test Set B. The clinical data consist of 40 patient records containing nurse validated vital signs data for heart rate (HR), systolic blood pressure (SBP), diastolic blood pressure (DBP), mean arterial pressure (MAP), and respiratory rate (RR) recorded at different intervals over a 10 hour period, but we had to drop one patient because they were missing respiratory rate values, thus only 39 patients were used for this experiment. Each instance is labeled as a member of one of two classes: 1) **H** for instances containing an AHE in the forecast window; and 2) **C** for instances containing no AHE in the forecast window. The forecast window was one hour after the last entry in the data set. Our approach is generically developed to predict a variety of clinical outcomes, not specifically AHE. The results in Table II demonstrate the efficacy of the TF, TF-IDF, and TF-IDF-IF on the Stacked BoP approach with this low frequency data. The highest classification results are in bold for each numerosity reduction level. The Stacked BoP, unlike the Multivariate BoP, can take advantage of numerosity reduction. Stacked BoP with TF-IDF-IF outperforms the raw data baseline and default accuracy of 65%. We also evaluated Stacked BoP with TF-IF, but the results consistently underperformed all other weightings.

	Raw Data	Bag-of-Patterns Data			
	Raw Time Series [†]	Stacked BoP	Stacked BoP TF	Stacked BoP TF-IDF	Stacked BoP TF-IDF-IF
NR=0	48.4%	100.0%	100.0%	99.7%	99.7%
NR=1	48.4%	99.4%	99.4%	97.4%	99.4%
NR=2	48.4%	96.3%	100.0%	96.3%	98.9%

TABLE I
1-NN CLASSIFICATION ACCURACY ON SYNTHETIC MULTIVARIATE TIME SERIES DATA WITH NO NUMEROSITY REDUCTION (NR=0), SOME NUMEROSITY REDUCTION (NR=1), AND ADVANCED NUMEROSITY REDUCTION (NR=2). [†]NUMEROSITY REDUCTION DOES NOT IMPACT THE RAW 1-NN CLASSIFICATION.

	Raw Data	Bag-of-Patterns Data			
	Raw Time Series [†]	Stacked BoP	Stacked BoP TF	Stacked BoP TF-IDF	Stacked BoP TF-IDF-IF
NR=0	56.4%	51.3%	48.7%	51.3%	43.6%
NR=1	56.4%	48.7%	64.1%	61.5%	66.7%
NR=2	56.4%	56.4%	53.9%	56.4%	71.8%

TABLE II
1-NN CLASSIFICATION ACCURACY ON HYPOTENSION CLINICAL TEST SET B WITH NO NUMEROSITY REDUCTION (NR=0), SOME NUMEROSITY REDUCTION (NR=1), AND ADVANCED NUMEROSITY REDUCTION (NR=2). [†]NUMEROSITY REDUCTION DOES NOT IMPACT THE RAW 1-NN CLASSIFICATION.

We upgraded the frequency of the test set by using the numeric data which is provided at the Physionet website. The numeric data is supplied by the monitors at the rate of one value per minute and is the middle frequency between the waveform data of the monitors (125hz) and the nurse validated data. At this frequency, the Stacked BoP representation was not able to run because of its large storage requirements at the alphabet sizes that produced the best results ($a > 12$). Thus, the remainder of our experiments are on the Multivariate BoP representation.

C. Multivariate Bag-of-Patterns

Using the high frequency data set for test set B, we were able to determine which were the best parameters for predicting hypotension in terms of SAX for the Multivariate BoP representation. After testing on window lengths of 60, 120 and 180 minutes, we found that using a window length of 120 minutes provided the best results. We also evaluated what number of symbols and alphabet size worked best. In all three window lengths, we discovered the number of symbols that corresponded to a symbol representing 3 minutes classified the greatest number of patients correctly. Thus, for a sliding window of 60 minutes, the number of symbols was 20; for 120, it was 40; and for 180, it was 60. The best alphabet size varied depending on which technique we used for classifying the data. The best results we were able to get for the numeric form of test set B are shown in Table III. We also decided to examine multiplying the term frequency by the probability of hypotension given an “MV word” (TF-P), which is the last category in the table.

Given that the best classification results for the greatest

number of techniques was determined when the sliding window (w) was size 120, the number of symbols (s) was 40 and the alphabet size (a) was 12, we used these parameters to determine which parameters best predicted hypotension using all combinations of 2, 3, 4 and 5 variables from the set of HR, SBP, DBP, MAP, and PP (Pulse Pressure, the delta of SBP and DBP). We found MAP and HR to be the best predictors of AHE as shown in Figure 6.

D. Univariate BoP versus Multivariate BoP

Our next set of experiments focused on comparing BoP using one variable and Multivariate BoP using two variables. We decided to use the Training Set from the 2009 Physionet Challenge for various reasons. The instances in the training set varied significantly in length from 2 to 24 days and included some gaps of several hours for a single patient. This data set contained the physiological data of 60 patients and had a default accuracy of 50% (30 with AHE in the forecast window, 15 with AHE outside the forecast window, 15 without AHE). Because two patients contained gaps of 3-4 days in their data, those patients were removed, and we used only 58 patients. Since both patients did not have AHE, the final data set had a default accuracy of 48%. Without the normalization of the natural language processing techniques, we expected the classification accuracy of BoP alone to suffer tremendously similar to text documents [23], but it did not. However, as seen in Table IV, our comparison of classification using the univariate BoP representation to the Multivariate BoP one gave the expected results. Multivariate BoP outperformed univariate BoP in all instances except the TF-P technique, which demonstrates the potential of multivariate time series

	Multivariate BoP	Multivariate BoP TF	Multivariate BoP TF-IDF	Multivariate BoP TF-IDF-IF	Multivariate BoP TF-IF	Multivariate BoP TF-P
a = 12	72.5%	77.5%	50.0%	77.5%	77.5%	75.0%
a = 13	70.0%	77.5%	45.0%	75.0%	75.0%	75.0%
a = 15	65.0%	70.0%	45.0%	75.0%	77.5%	70.0%
a = 16	70.0%	77.5%	47.5%	75.0%	72.5%	75.0%
a = 17	70.0%	77.5%	50.0%	75.0%	75.0%	77.5%

TABLE III

1-NN CLASSIFICATION ACCURACY ON HIGH FREQUENCY HYPOTENSION TEST SET B USING HR AND MAP USING $w = 120$, $s = 40$. a REPRESENTS THE ALPHABET SIZE.

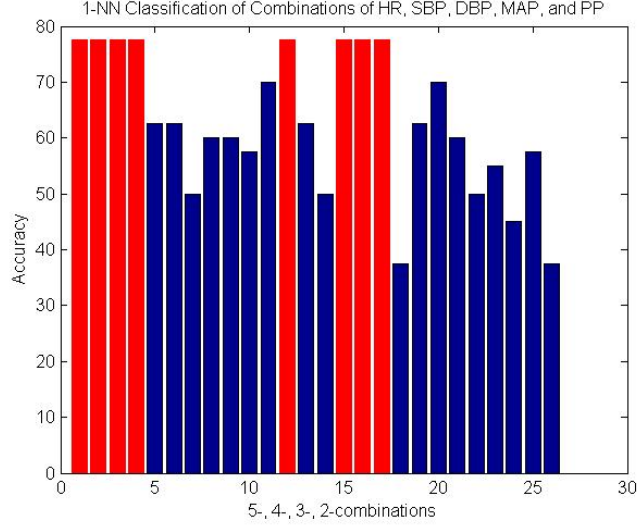


Fig. 6. Combinations that included HR and MAP are in red. Calculated for TF. Similar results with all techniques, except Multivariate TF-IDF ($w = 120$ $s = 40$ $a = 12$)

	BOP	TF	TF-IDF	TF-IDF-IF	TF-IF	TF-P
HR	58.62%	55.17%	44.83%	58.62%	60.34%	58.62%
MAP	56.90%	56.90%	55.17%	56.90%	55.17%	56.90%
HR, MAP	63.79%	62.07%	60.34%	63.79%	63.79%	56.9%

TABLE IV

1-NN CLASSIFICATION USING UNIVARIATE BoP VERSUS MULTIVARIATE BoP ON THE HYPOTENSION TRAINING SET USING $w = 120$, $s = 30$, AND $a = 15$. THE MULTIVARIATE REPRESENTATION OUTPERFORMED THE UNIVARIATE IN ALL CASES EXCEPT TF-P.

analysis in patient classification.

V. CONCLUSIONS AND FUTURE WORK

Submitted results from the PhysioNet/Computers in Cardiology Challenge 2009: Predicting Acute Hypotensive Episodes challenge outperform our approach. For example, AHE-tailored approaches using neural networks [28], rule-based systems [29], and feature engineering [30] produce classification accuracies of 92.5%, 87.5%, and 70%, respectively. However, these approaches were developed on the 60 patient training set for the specific purpose of predicting AHE whereas our approach is designed for any highly dependent multivariate time series data set.

We presented an extension to the univariate BoP representation for classification of multivariate time series data. Furthermore, our approach incorporates weighted frequency counts using term frequency, inverse document frequency, and inverse frequency. Preliminary results on a multivariate vital signs dataset from a challenge task to predict AHE demonstrate the potential of our multivariate time series representations. Our results outperform the baseline 1-NN classification accuracy, but do not exceed results tailored to predicting AHE.

Future work will proceed in three directions. First, we will explore other techniques that capture temporal interdependencies between vital signs using Multivariate Bag-of-Patterns such as probability and support vector machines using linear,

polynomial and radial basis function kernels. Second, we will apply our techniques to other domains that have similar characteristics, such as applying the Stacked Bags-of-Patterns to robot sensor data. Finally, our Multivariate Bag-of-Patterns representation is scalable to high frequency vital signs data. We will apply our methods to additional medical datasets for predicting various patient outcomes.

REFERENCES

- [1] T. Heldt, W. Long, G. C. Verghese, P. Szolovits, and R. G. Mark, "Integrating data, models, and reasoning in critical care," *Proceedings of the 28th IEEE EMBS International Conference*, vol. 28, pp. 350–353, September 2006.
- [2] G. S. Halford, R. Baker, J. McCredden, and J. D. Bain, "How many variables can humans process?" *American Psychological Society*, vol. 16, no. 1, 2005.
- [3] P. Ordóñez and M. desJardins, "Multivariate time series analysis of clinical and physiological data," in *Proceedings of Grace Hopper Celebration of Women in Computing*, 2010, pp. 31–36.
- [4] B. S. Harry Hochheiser, "Dynamic query tools for time series data sets, timebox widgets for interactive exploration," *Information Visualization* 3, pp. 1–18, March 2004.
- [5] S. Lam, "Patternfinder in microsoft amalg: Temporal query formulation and result visualization in action," Master's thesis, University of Maryland College Park, 2008.
- [6] T. D. Wang, C. Plaisant, A. J. Quinn, R. Stanchak, S. Murphy, and B. Shneiderman, "Aligning temporal data by sentinel events: discovering patterns in electronic health records," in *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, ser. CHI '08. New York, NY, USA: ACM, 2008, pp. 457–466. [Online]. Available: <http://doi.acm.org/10.1145/1357054.1357129>
- [7] M. Saeed and R. Mark, "A novel method for the efficient retrieval of similar multiparameter physiologic time series using wavelet-based symbolic representations," in *AMIA 2006 Symposium Proc.*, 2006, pp. 679–683.
- [8] X. Ge and P. Smyth, "Deformable markov model templates for time-series pattern matching," in *Sixth International Conference on Knowledge Discovery and Data Mining*, 2000, pp. 81–90.
- [9] E. J. Keogh, K. Chakrabarti, M. J. Pazzani, and S. Mehrotra, "Dimensionality reduction for fast similarity search in large time series databases," *Journal of Knowledge and Information Systems*, vol. 3, no. 3, pp. 263–286, 2001.
- [10] J. Himberg, K. Korpiaho, H. Mannila, J. Tikanmäki, and H. Toivonen, "Time series segmentation for context recognition in mobile devices," in *ICDM '01: Proceedings of the 2001 IEEE International Conference on Data Mining*. Washington, DC, USA: IEEE Computer Society, 2001, pp. 203–210.
- [11] J. Catalano, T. Armstrong, and T. Oates, "Discovering patterns in real-valued time series," in *Proceedings of the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD)*, 2006.
- [12] E. Keogh, "A decade of progress in indexing and mining large time series databases," in *Proceedings of the 32nd international conference on Very large data bases*, ser. VLDB '06, 2006, pp. 1268–1268.
- [13] B. Chiu, E. Keogh, and S. Lonardi, "Probabilistic discovery of time series motifs," in *9th International Conference on Knowledge Discovery and Data Mining (SIGKDD'03)*, 2003, pp. 493–498.
- [14] D. Yankov, E. Keogh, J. Medina, B. Chiu, and V. Zordan, "Detecting time series motifs under uniform scaling," in *KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM, 2007, pp. 844–853.
- [15] D. Minnen, C. Isbell, I. Essa, and T. Starner, "Detecting subdimensional motifs: An efficient algorithm for generalized multivariate pattern discovery," in *IEEE Int. Conf. on Data Mining (ICDM)*, vol. 1, 2007.
- [16] K. pong Chan and A. W.-C. Fu, "Efficient time series matching by wavelets," in *Fifteenth International Conference on Data Engineering*, 1999, pp. 126–133.
- [17] R. Agrawal, C. Faloutsos, and A. N. Swami, "Efficient similarity search in sequence databases," in *Proceedings of the 4th International Conference on Foundations of Data Organization and Algorithms*, ser. FODO '93. London, UK: Springer-Verlag, 1993, pp. 69–84.
- [18] K. Chan and A. Fu, "Efficient time series matching by wavelets," in *Proceedings of the 15th International Conference on Data Engineering*, ser. ICDE '99. Washington, DC, USA: IEEE Computer Society, 1999, pp. 126–.
- [19] J. Lin, E. Keogh, L. Wei, and S. Lonardi, "Experiencing sax: a novel symbolic representation of time series," *Data Min. Knowl. Discov.*, vol. 15, pp. 107–144, October 2007.
- [20] J. Shieh and E. Keogh, "isax: indexing and mining terabyte sized time series," in *KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM, 2008, pp. 623–631.
- [21] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *DMKD '03: Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*. New York, NY, USA: ACM Press, 2003, pp. 2–11.
- [22] J. Lin and Y. Li, "Finding structural similarity in time series data using bag-of-patterns representation," in *Proceedings of the 21st International Conference on Scientific and Statistical Database Management*, ser. SSDBM 2009. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 461–477.
- [23] G. Salton, *Automatic text processing: the transformation, analysis, and retrieval of information by computer*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1989.
- [24] Y. Morinaka, M. Yoshikawa, T. Amagasa, and S. Uemura, "The 1-index: An indexing structure for efficient subsequence matching in time sequence databases," in *Proceedings of PAKDD*, 2001, pp. 51–60.
- [25] E. Keogh and T. Folias, "The ucr time series data mining archive," 2002. [Online]. Available: http://www.cs.ucr.edu/~eamonn/time_series_data/
- [26] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [27] —, "Physionet/computers in cardiology challenge 2009: Predicting acute hypotensive episodes," 2009. [Online]. Available: <http://www.physionet.org/challenge/2009/>
- [28] J. Henriques and T. Rocha, "Prediction of acute hypotensive episodes using neural network multi-models," in *Computers in Cardiology*, 2009. IEEE, 2009, pp. 549–552.
- [29] M. Mneimneh and R. Povinelli, "A rule-based approach for the prediction of acute hypotensive episodes," in *Computers in Cardiology*, 2009. IEEE, 2009, pp. 557–560.
- [30] P. Langley, S. King, D. Zheng, E. Bowers, K. Wang, J. Allen, and A. Murray, "Predicting acute hypotensive episodes from mean arterial pressure," in *Computers in Cardiology*, 2009. IEEE, 2009, pp. 553–556.