Multivariate Methods for Classifying Physiological Data

Patricia Ordóñez* Tom Armstrong† Tim Oates‡ Jim Fackler§
Christoph U. Lehmann¶

Abstract

In this paper we examine two novel multivariate time series representations to classify physiological data of different lengths, *Multivariate Bag-of-Patterns* and *Stacked Bags-of-Patterns*. We also borrow techniques from the natural language processing and text mining (e.g., term frequency and inverse document frequency) to improve classification accuracy. We compare how these multivariate representations classify the data compared to extensions of two univariate representations, known as Piecewise Dynamic Time Warping and Bag-of-Patterns, into the multivariate domain. We present experimental results on classifying adult patients who have experienced acute episodes of hypotension (AHE) and neonatal patients who have experienced a patent ductus arteriosus (PDA). We also evaluated how these methods fared in classifying robotic sensor data to determine location and direction of the robot and motion capture data to differentiate types of motions to determine whether the methods are generalizable to other domains.

1 Introduction

Much of the previous research on time series analysis has focused on univariate time series data [1]. Our interest is in examining multiple time series per individual where each time series may be of different lengths and the ranges for each variable may differ by orders of magnitude.

Current methods for measuring the well being of a patient in the ICU acquire a patient's vital signs data at rates that are difficult for a human to analyze (1-500 Hertz). These measurements are displayed on a monitor for a few seconds and then lost to further analysis. Instead, a lower-frequency version of this data is stored in an electronic health record after validation by a medical provider at the rate of once every 15 minutes to once every several hours, and physicians make life-saving decision based on this lower-frequency data. Recently, however, there has been interest in storing and analyzing the high-frequency data using automated and semi-automated methods [2].

1.1 The Challenges of Working with Medical Data

Vital signs such as heart rate, laboratory data such as blood sugar levels, and clinical data such as urine output, individually fall under the category of univariate time series data. Together, however, they measure the overall state of a patient. Thus, physiological and clinical data are multivariate time series data. However, there are no standards in terms of which parameters should be used for measuring health. Parameters are selected based on the provider's clinical instinct and expertise. Therefore, the available parameters for analysis vary greatly from one individual patient to another, and from one period of time to another. Therein lies one of the challenges in comparing this high-frequency data from one patient to another, the lack of a uniform standard for collecting physiological and clinical data.

Another challenge is storing the data for comparison. Since so much data is being generated, representations that are compressed yet maintain enough information about the original data that inferences can be made about the data are critical.

One of the objectives of our research is to prove that multivariate analysis on high-frequency data using machine learning techniques can be effective at grouping similar patients, even if only a few parameters are used. A second objective is to create a representation that makes it easy to store and analyze the data with the goal of eventually developing similarity-based retrieval methods for patients.

In the following sections, we review related work on time series discretization methods. Next, we discuss the representations we have created, Multivariate Bag-of-Patterns and Stacked Bags-of-Patterns, for classifying multivariate time series data. Then, we present some of the natural language processing/text mining techniques we incorporated in the classification of the data, including one technique inspired by the work of Saeed and Mark [3]. Finally, we conclude with empirical results of the Multivariate Bag- and Stacked Bags-of-Patterns representations on multivariate data (e.g., physiological

*University of Puerto Rico Río Piedras Campus
†Wheaton College (MA)
‡University of Maryland Baltimore County
§Johns Hopkins School of Medicine
¶Vanderbilt University

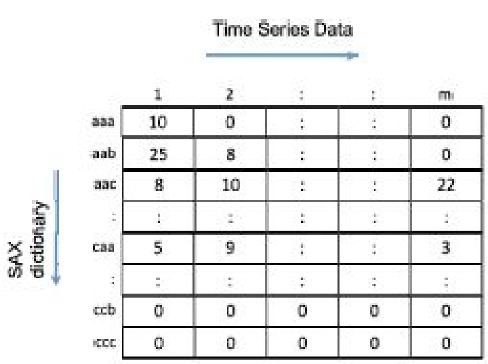
data, robotic sensor data, and motion capture data).

2 Background and Related Work

This work expands on previous work. We will review the bare essentials of the novel multivariate representations here, but for a more through explanation of the Multivariate Bag-of-Patterns and Stacked-Bags-of-Patterns representations, please see previous work [4].

2.1 Bag-of-Patterns (BoP) Bag-of-Patterns [5] outperforms existing techniques for classification, clustering and anomaly detection of univariate time series. The BoP representation takes the Symbolic Aggregate ApproXimation [6] representation of 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 two documents would be compared in a bag-of-words style, by examining the histograms of their word frequencies.

In BoP, each univariate time series is converted into a frequency vector. The vectors are created by first converting the time series into its SAX representation. SAX uses a window of length w to capture patterns in small local subsequences and convert the subsequences into a SAX representation (i.e., a word), and then slides across the time series to capture all subsequences in a window of length w. Identical consecutive words are only counted once when numerosity reduction is set as the default in SAX, to avoid matching based on recurring patterns (as with stop words in a document). In BoP, the frequency of the SAX words is stored in the BoP vector. Figure 1 displays the frequency vectors for a collection of m time series. The distance between the vectors can be calculated using any conventional distance metric, like Euclidean distance.



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

Figure 1: BoP representations of m time series [5]

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 1, the BoP representations for m time series are displayed in a matrix.

3 Multivariate Representations of Bag-of-Patterns

We expand on the BoP representation by tailoring it to multivariate medical data, where emphasizing specific patterns that represent a clinical disease state is critical to classifying the patients correctly. For this purpose, we incorporate techniques from natural language processing/text mining [7] and apply them to the multivariate BoP representations.

3.1 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 with organ interdependency.

Multivariate BoP takes multiple univariate times series and converts them into their SAX representation individually. There is no numerosity reduction, so the SAX representations for the different parameters are the same length for a given patient. The trends across the parameters in time are represented by creating multivariate words, "MV words," that represent the state of the patient within one single interval. So an "MV word" is a single time step across multiple variables instead of a window of a single variable. Once the "MV words" are created, the Bag-of-Patterns approach is applied to create a frequency vector for the patient. See Figure 2 for a visualization of processing a single patient's data with two variables.

3.2 Stacked Bags-of-Patterns The second multivariate Bag-of-Patterns representation is *Stacked Bags-of-Patterns* (Stacked BoP). As in Mulitvariate 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 3

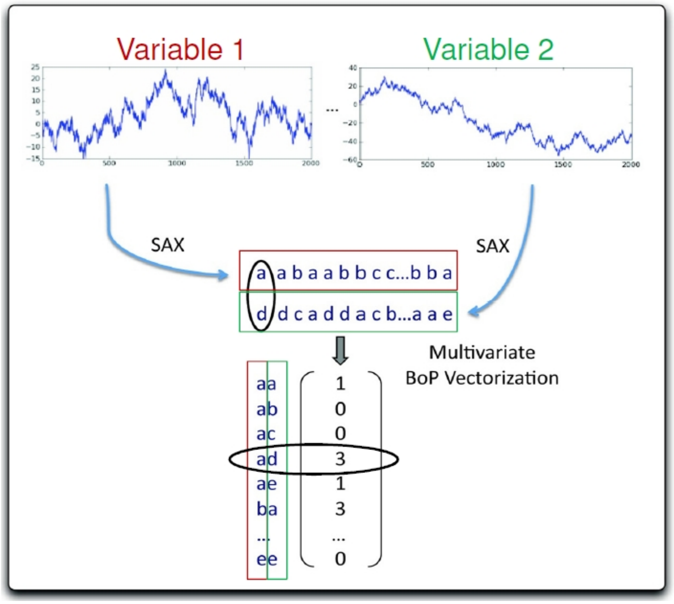


Figure 2: Multivariate Bag-of-Patterns for a single 2-variable instance

for a visualization of processing a single instance with n variables.

3.3 Adapted Natural Language Processing/Text Mining Techniques The univariate BoP approach has proven effective with time series of the same length. However, the current literature does not test BoP with time series of different lengths, as most patient vital signs data are collected. We have incorporated several language processing techniques: term frequency (TF), inverse document frequency (IDF), and inverse frequency (IF) to improve classification anticipating that BoP would not work as well on time series of substantially different lengths. A combination of the last two techniques was implemented in IDF-IF. In the formulas, x_i is a "MV" or "SAX" word, n is the number of instances (e.g., patients), and m is the number of words in the vocabulary over the SAX alphabet, a^s. The number of instances that contain the word x_i is $f_n(x_i)$, $f_a(x_i)$ is the total number of occurrences of the word x_i in all instances, and $f_j(x_i)$ is the frequency of x_i in the j^{th} instance. Each of the multivariate BoP vectors for a single instance is weighted term by term using combinations of the following such as TF-IDF, TF-IF, and TF-IDF-IF:

$$\mathtt{TF}(x_i) = \frac{f_j(x_i)}{\sum_{i=1}^{m} f_j(x_i)}$$

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

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

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

Bag-of-Patterns (BoP) is based on the vector space model [8] (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 is a method for giving more weight to unusual words, so that the similarity between two documents is measured by the number of uncommon words they have in common.

Both IF and IDF-IF were developed based on the work of Saeed and Mark [3]. The multivariate BoP approach allows the user to perform patient "word frequency" comparisons similar to those that are performed on documents. The idea 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 state, since these are not the normal states of the patient population.

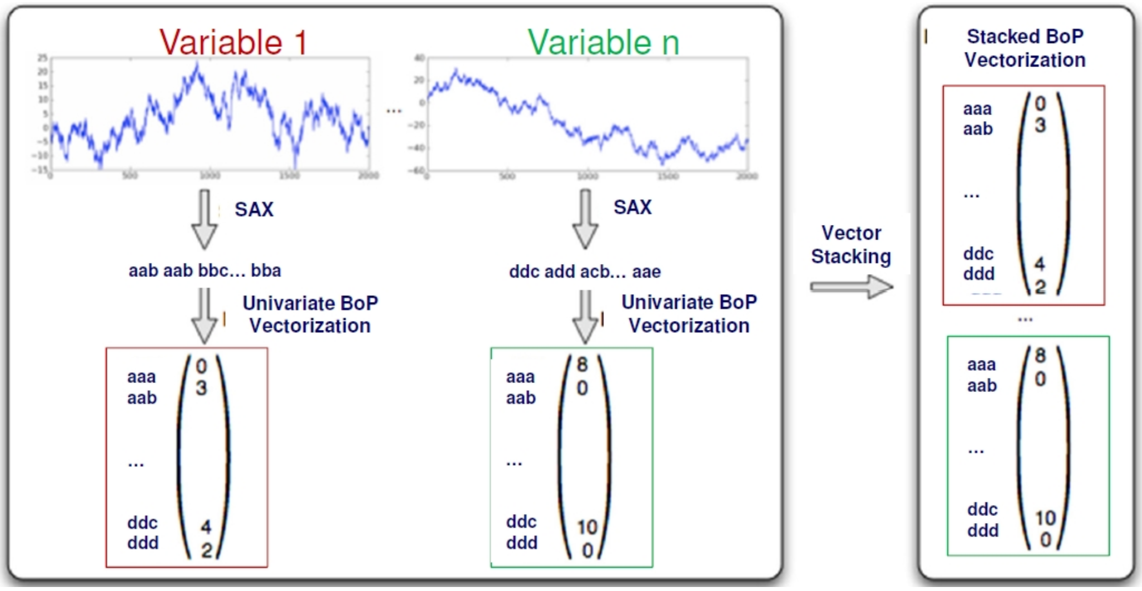


Figure 3: Stacked Bags-of-Patterns for a single n-variable instance

4 Evaluation Method

The aims of the final evaluation were (1) to compare the precision of multiple univariate Bag-of-Pattern representations with that of two multivariate Bag-of-Patterns representations in the classification of multivariate time series data, (2) to compare how the latter representations performed relative to other multivariate representations, and (3) to examine how these methods fared in three domains: medicine, robotics, and motion capture.

4.1 Objectives

- **Aim 1:** We hypothesize that our multivariate Bag-of-Patterns representations and methods would classify multivariate time series data more accurately than the univariate Bag-of-Patterns representations.

- **Aim 2:** We hypothesize that our multivariate Bag-of-Patterns representations and methods will outperform current multivariate methods for classifying multivariate time series data [9]).

- **Aim 3:** We hypothesize that our multivariate Bag-of-Patterns representations and methods will generalize to other multivariate time series domains.

4.2 Data Sets All of our experiments were performed on six real-world data sets. Two of the data sets were medical data sets. The first of these came from the Physionet Challenge 2009[1] and consisted of 1-6 days of high-frequency physiological data from patients in a Critical Care Unit. The classification task in this data set was to determine which patients were going to enter an episode of acute hypotension in the forecast window of one hour after the last entry in the data set. Because in our preliminary results only HR and MAP were determined to be good predictors of hypotension, only those two variables were used.

The second consisted of HL7 high-frequency neonatal patient data from the Johns Hopkins Neonatal Intensive Care Unit monitors used in the previous work for the evaluation of the visualization. The classification task in this case was to determine which patients were experiencing a period of Patent Ductus Arteriosus (PDA). However, because a Pediatric Cardiologist was not available to label the data by examining echocardiograms, the gold standard for diagnosing PDA, this data set was dropped. Instead we used the low-frequency nurse-validated data set (recorded every 15 minutes to hourly) which was used to extract the eight patients that were used in the visualization for the Fixed Dual MTSA Visualization [10]. This data consisted of 91 neonatal patients, 40 of which had PDA. The data did have large regions of missing data which had to be interpolated over 30 minute intervals before use. Five variables were used in the evaluation as recommended by the physicians. These variables were selected because: they had

[1]This data was obtained from http://www.physionet.org/challenge/2009/.

the least amount of missing data; they were the most common variables examined by physicians in the evaluation (in the case of pulse pressure, physicians derived pulse pressure when using the tables in the evaluation of the visualization); or as in the case of heart rate (HR) and mean arterial pressure (MAP), they were indicated by the preliminary tests as the best indicators of hypotension. These variables were HR, MAP, diastolic blood pressure (DBP), pulse pressure (PP), and respiratory rate (RR).

The next two data sets were robotic sensor data from the Wheaton College Autonomous Learning Lab. The first data set consisted of sonar readings from the foot of a MobileRobots PeopleBot, as seen in Figure 4.2. The data was collected in 10-20 second intervals as the robot was asked to maneuver through a maze. Four different maze configurations were used. The data was used to determine whether the robot was moving through a straightaway or going around a corner. The robot provided sensor values from 24 sensors and 1 bumper. For the experiments, only 5 sensor values were used from the front base of the machine; two on the sides and three spread along the front of the machine. Each variable contained between 130-350 data points.

The second set of robotic sensor data was gathered while the robot was stationary and people walked across its path from left to right or right to left. The classification task for this data set was to determine the direction of the person crossing the robot's path. This data set was the smallest data set containing instances have between 70-100 data points.

The final two data sets were motion capture data sets. The first belongs to the Carnegie Mellon University Graphics Lab Motion Capture Database.[2] The data set consisted of two motions – jumping and running. The sensors were placed on the people's skin. The location of the sensors were on the front and back sides of the body and the hand and foot. Figure 4.2 shows where some of the sensors were placed on the body. The experiments were run using only 5 variables from the hand, which ranged from 130 to 1500 points per subject.

The second motion capture data set is the Physical Activity Monitoring for Aging People data set. The data set contains 8 subjects performing 14 activities. Sensors were placed on the person's body as they were performing activities indoors and outdoors. Indoor activities included ironing, climbing stairs, and vacuuming; outdoor activities included running, cycling, and jumping rope. The classification task was to determine whether the person was performing indoor or outdoor activities. This data set was the largest of all the data sets with 70,000 to 220,000 of points per subject. While 62 different variables were available, only five variables for the hand were used assuming these values would be the most differentiating.

[2] The data was obtained from mocap.cs.cmu.edu. The database was created with funding from NSF EIA-0196217.

4.3 Baselines We will be evaluating the data using one baseline, which is the default accuracy where all instances are classified as the majority class. Then we will compare the multivariate representations to two multivariate interpretations of univariate representations. The first is *Ensemble Voting with Bag-of-Patterns*. The second is *Multivariate Piecewise Dynamic Time Warping*.

4.3.1 Ensemble Voting with Bag-of-Patterns In Ensemble Voting, each variable for an instance is converted into its BoP representation. Using 1 Nearest Neighbor with Euclidean distance, each BoP representation is used to classify the instances in the database. To make the approach multivariate, each instance is classified by using the outcome of the univariate classifications to vote on the final label for the instance. Whereas ensemble voting is well known machine learning technique for classification, combining ensemble voting with Bag-of-Patterns is our attempt at creating a multivariate approach that emphasizes univariate patterns.

4.3.2 Piecewise Dynamic Time Warping Dynamic Time Warping uses dynamic programming to find a minimum cost of mapping one time series to another. Cost is determined by calculating the distance between two points in time using Euclidean distance. In the case of Piecewise Dynamic Time Warping (PDTW) [9], the cost is usually the square of difference between the PAA segments. For the multivariate Piecewise Aggregate Approximation implementation that we used, the cost is the Euclidean distance between vectors of the PAA representations across all the variables at a point in time. To our knowledge, this approach is the first application of PDTW to multivariate time series data.

5 Experimental Results

The results for accuracy, precision, recall and F-measure are displayed in Tables 1, 2, 3, and 4, respectively. The results for applying NLP/Texting Mining-inspired techniques to Stacked Bags-of-Patterns and Multivariate Bag-of-Patterns representations are found in Tables 5 and 6, respectively. These results were gathered by using the highest F-measure as the standard for the selection of the best results. Thus, the values for accuracy, precision, and recall do not refer to the highest values that were obtained in each category, but rather to the

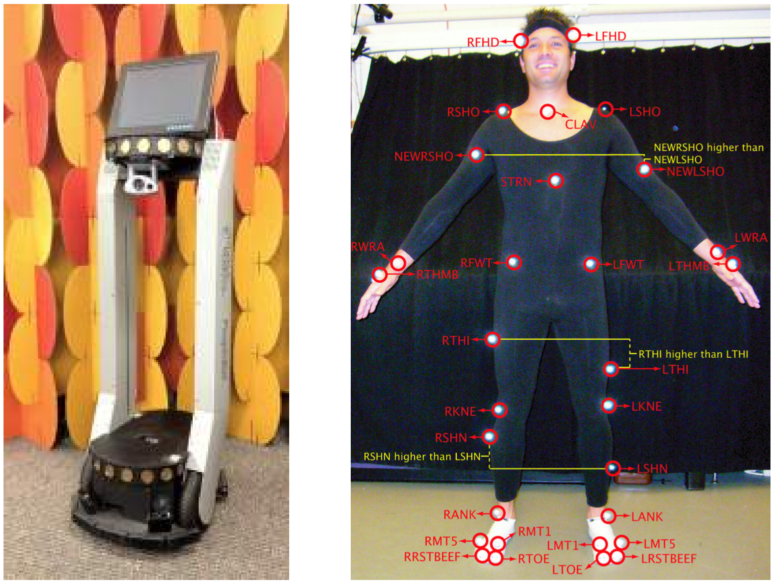


Figure 4: The MobileRobots PeopleBot from the Wheaton College Autonomous Learning Lab on left and Sensor placement of the CMU Mocap data set on right

values that correspond to the highest value of F-measure displayed in Table 4. The values were calculated using the following equations:

$$Acc. = \frac{truePos + trueNegs}{\#instances}$$

$$Prec. = \frac{truePos}{truePos + falsePos}$$

$$Recall = \frac{truePos}{truePos + falseNegs}$$

$$Fmeas. = \frac{2 * Prec. * Recall}{Prec. + Recall}$$

6 Discussion

The first of the three hypothesis was that the multivariate Bag of Patterns representations would do better than univariate Bag-of-Patterns representations on multiple variables. We implemented Ensemble Voting as the classifier for the univariate representation to test this hypothesis. One Nearest Neighbor on each variable's Bag-of-Patterns representation was used as the first level classifier, and then each classifier became a vote for the final classifier. The second hypothesis was that multivariate Bag-of-Patterns representations and methods will outperform current multivariate methods for classifying multivariate time series data. We implemented a multivariate version of Piecewise Dynamic Time Warping to test that hypothesis. Because in medicine, the two primary evaluation metrics are precision (better known as Positive Predicative Value) and recall (better known as Sensitivity), we focused on F-measure, which is a metric that approaches one when both precision and recall are high. For the third hypothesis, we tested four additional multivariate time series data sets using these parameters.

In the medical domain, the best F-measure results were using PDTW for the hypotension data set and Ensemble Voting for the PDA data. Considering that the data for the hypotension data varied from one day in length to six days in length, the result for PDTW is surprising (0.841); however, the multivariate bag-of-patterns was not far behind (0.77). The F-measure for the Ensemble Voting technique for the PDA was 0.675 was not that far off from that of Stacked BoP or Multivariate Bop, 0.625 and 0.632, respectively. In both cases, the Multivariate BoP performed better that Stacked BoP. However, our hypothesis was that the multivariate representations would outperform the univariate ones and current methods the classification of medical data and it did not.

Another observation is that in terms of variability, PDTW gave more consistent results. For example, for 18 runs with compression ratios ranging from 15-100 in increments of 5 on the hypotension data set, the average F-measure was 0.8218 and the standard deviation was 0.0099, whereas the average F-measure and standard deviation for Ensemble Voting, Multivariate Bag-of-Patterns and Stacked Bags-of-Patterns were

	Baseline	Univariate Ext		Multivariate BoP	
	Default	E Voting†	PDTW	StBoP	MvBoP
Hypotension (58)	0.483	0.638	**0.81**	0.64	0.76
PDA (91)	0.560	**0.714**	0.692	0.67	0.692
Robot Corner (111)	0.505	**0.982**	0.883	0.964	0.91
Robot L2R (50)	0.5	0.8	**0.94**	0.76	0.7
Mocap jump(40)	0.5	**0.975**	0.925	**0.975**	**0.975**
PAMAP (15) indoors	0.533	**1.0**	0.667	**1.0**	**1.0**

Table 1: 1-NN classification **Accuracy** on various data sets based on best F-measure values. The default accuracy was based on classifying everyone as healthy for the medical data sets and on classifying everyone as belonging to the majority class for the remaining procedures. The parenthetical value for each data set is the number of instances in the data set. †Ensemble Voting.

	Baseline	Univariate Ext		Multivariate BoP	
	Default	E Voting†	PDTW	StBoP	MvBoP
Hypotension (58)	0.517	0.68	0.744	**0.79**	0.78
PDA (91)	0.440	0.675	**0.929**	0.625	0.667
Robot Corner (111)	0.505	0.982	**1.0**	0.948	0.976
Robot L2R (50)	0.5	0.759	**1.0**	0.71	0.696
Mocap jump (40)	0.5	1.0	1.0	1.0	1.0
PAMAP indoors (15)	0.534	1.0	1.0	1.0	1.0

Table 2: 1-NN classification **Precision** on various data sets based on best F-measure values. The parenthetical value for each data set is the number of instances in the data set. Default is the proportion of instances belonging to the target class. †Ensemble Voting.

	Baseline	Univariate Ext		Multivariate BoP	
	Default	E Voting†	PDTW	StBoP	MvBoP
Hypotension (58)	0.517	0.567	**0.967**	0.73	0.8
PDA (91)	0.440	**0.675**	0.325	0.625	0.6
Robot Corner (111)	0.505	0.982	0.768	**1.0**	0.976
Robot L2R (50)	0.5	**1.0**	0.88	0.92	0.88
Mocap jump (40)	0.5	**0.95**	0.85	**0.95**	**0.95**
PAMAP indoors (15)	0.534	**1.0**	0.375	**1.0**	**1.0**

Table 3: 1-NN classification **Recall** on various data sets based on best F-measure values. The parenthetical value for each data set is the number of instances in the data set. Default is the proportion of instances belonging to the target class. †Ensemble Voting.

	Baseline	Univariate Ext		Multivariate BoP	
	Default	E Voting†	PDTW	StBoP	MvBoP
Hypotension (58)	0.517	0.618	**0.841**	0.68	0.77
PDA (91)	0.440	**0.675**	0.418	0.625	0.632
Robot Corner (111)	0.505	**0.982**	0.869	0.966	0.912
Robot L2R (50)	0.5	0.815	**0.936**	0.786	0.727
Mocap jump (40)	0.5	0.974	0.919	0.974	0.974
PAMAP indoors (15)	0.534	**1.0**	0.545	**1.0**	**1.0**

Table 4: 1-NN classification **F-Measure** on various data sets. The parenthetical value for each data set is the number of instances in the data set. Default is the proportion of instances belonging to the target class. †Ensemble Voting.

	Stacked Bags-of-Patterns				
	StBoP	TF	TFIDF	TFIDFIF	TFIF
Hypotension (30/58)	0.68	0.75	0.7	0.75	0.7
PDA (40/91)	0.625	0.581	0.617	0.581	0.581
Robot Corner (56/111)	0.966	0.949	0.947	0.957	0.957
Robot L2R (25/50)	0.768	0.807	0.755	0.767	0.733
Mocap jump (20/40)	0.974	0.952	0.816	**0.976**	**0.976**
PAMAP (8/15) indoors	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**

Table 5: 1-NN classification **F-Measure** on various data sets using NLP/Texting Mining-inspired techniques on Stacked Bags-of-Patterns representations. Value in parenthesis is ratio of instances in the target class of the data set.

	Multivariate Bags-of-Patterns				
	BoP	TF	TFIDF	TFIDFIF	TFIF
Hypotension (30/58)(58)	0.77	0.68	0.71	0.7	0.7
PDA (40/91)	0.632	0.625	0.625	0.627	0.634
Robot Corner (56/111)	0.912	0.917	0.906	0.924	0.887
Robot L2R (25/50)	0.727	0.727	0.667	0.767	0.733
Mocap jump (20/40)	0.974	0.927	0.76	0.95	0.95
PAMAP indoors (8/15)	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**

Table 6: 1-NN classification **F-Measure** on various data sets using NLP/Texting Mining-inspired techniques on Multivariate Bag-of-Patterns representations. Value in parenthesis is ratio of instances in the target class of the data set.

0.6156±0.2428, 0.7073±0.0343, and 0.4811±0.1774, respectively. Therefore, until there is a way to extract the best values for the four parameters for SAX, these methods are too variable to be applicable on large scale data. PDTW, on the other hand, has only one parameter which is the compression ratio.

However, when we moved to working on the very large motion capture data, the PAMAP data set, determining the correct compression value was very difficult. We did a compression ration from 100 to 2000 increasing by a hundred on each run. The average F-measure for those runs was 0.48±0.133. Some subjects contained 72,000 data points where others contained 220,000 data points. In these cases, the methods that used a Bag-of-Pattern representation had perfect accuracy. While definitive conclusions can not be made, it merits further investigation to determine whether these representations are better able to handle large data sets than PDTW. One advantage to using these representation on larger data sets is that the algorithms to create these representations are amenable to parallel processing.

The results in other domains were promising which brings us to our final hypothesis. Once again either PDTW or Ensemble Voting outperformed the multivariate Bag-of-Patterns methods, rejecting our first and second hypothesis. We tested both multivariate Bag-of-Patterns representation by incorporating the natural language processing/text mining techniques discussed earlier (TF, TFIDF, TFIDFIF, and TFIF) as seen in Tables 5 and 6, but the techniques did not improve F-measure values to the point of surpassing the Ensemble Voting and/or PDTW techniques except for with the Motion Capture data. However, our final hypothesis was supported and that is that these methods are generalizable to other domains.

7 Conclusion

In this paper, we compared how two multivariate time series representations, Multivariate Bag-of-Patterns and Stacked-Bags-of-Patterns, compared to two multivariate representations from the univariate domain, Multivariate PDTW and Ensemble Voting BoP. We evaluated these representations on 6 multivariate time series data sets from three different domains: medicine, robotics, and motion capture. Future work will include investigating using our representations in the classification of multivariate domain using few variables.

References

[1] 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.

[2] G. D. Clifford, W. J. Long, G. B. Moody, and P. Szolovits, “Robust parameter extraction for decision support using multimodal intensive care data,” *Philosophical Transactions of the Royal Society - Series A: Mathematical, Physical and Engineering Sciences*, vol. 367, no. 1887, pp. 411–429, 2009.

[3] M. Saeed and R. Mark, “A novel method for the efficient retrieval of similar multiparameter physiologic time series using wavelet-based symbolic representations,” in *Proceedings of the AMIA 2006 Annual Symposium*, 2006, pp. 679–683.

[4] P. Ordóñez, T. Armstrong, T. Oates, and J. Fackler, “Classification of patients using novel multivariate time series representations of physiological data,” in *International Conference of Data Mining in Workshop of Mining Multiple Sources of Information, to appear*, 2011.

[5] 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.

[6] J. Lin, E. Keogh, L. Wei, and S. Lonardi, “Experiencing SAX: a novel symbolic representation of time series,” *Data Min Knowl Disc*, vol. 15, pp. 107–144, 2007.

[7] G. Salton, *Automatic text processing: the transformation, analysis, and retrieval of information by computer*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1989.

[8] K. Chan and A. Fu, “The l-index: An indexing structure for efficient subsequence matching in the sequence databases,” in *Proceedings of the Fifth Pacific Asia Conference on Knowledge Discovery and Data*, 2001, pp. 51–60.

[9] E. J. Keogh and M. J. Pazzani, “Scaling up dynamic time warping for datamining applications,” *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining KDD 00*, pp. 285–289, 2000.

[10] P. Ordóñez, T. Oates, M. Lombardi, G. Hernández, K. W. Holmes, J. Fackler, and C. U. Lehmann, “Visualization of multivariate time series data in a neonatal icu,” *IBM Journal of Research and Development, to appear*, 2012.