

Sensor Based Time Series Classification of Body Movement

Swapna Philip¹, Yu Cao^{2,*}, and Ming Li¹

¹ Department of Computer Science
California State University, Fresno
Fresno, CA, U.S.A.

swapna.philip@gmail.com, mingli@csufresno.edu

² College of Engineering and Computer Science
University of Tennessee at Chattanooga
Chattanooga, TN, U.S.A.
yu-cao@utc.edu

Abstract. Advances in sensing and monitoring technology are being incorporated into today's healthcare practice. As a result, the concept of Body Sensor Networks (BSN) has been proposed to describe the wearable/wireless devices for healthcare applications. One of the major application scenarios for BSN is to detect and classify the body movements for long-term lifestyle and healthcare monitoring. This paper introduces a new approach for analyzing the time series obtained from BSN. In our research, the BSN record the acceleration data of the volunteer's movement while performing a set of activities such as jogging, walking, resting, and transitional activities. The main contribution of this paper is the proposed time series approximation and feature extraction algorithm that can convert the sensor-based time series data into a density map. We have performed extensive experiments to compare the accuracy in classifying the time series into different activities. It is concluded that the proposed approach would aid greatly the development of efficient health monitoring systems in the future. To the best of our knowledge, no similar research has been reported in the BSN field and we expect our research could provide useful insights for further investigation.

Keywords: Body Sensor Networks, Time Series Analysis, Human body movement, machine learning, classification.

1 Introduction

Body sensor networks (BSN) [1, 2] enable an inexpensive approach for data collection. The BSN include sensor nodes that communicate wirelessly and they function with less consumption of power and less memory capacity. These sensor networks are very useful especially in the field of health monitoring as they are capable of monitoring temperature, humidity, force, acceleration and heartbeat [3, 4]. The sensor networks can

* Corresponding author.

thus be specifically used in interpreting human motion and thus in health care for the elderly. The BSN have their own advantage over other types of motion or health monitoring systems as they are wearable, less expensive and more accurate with negligible environmental disturbances affecting data collection.

While monitoring the health of patients, it is necessary to deal with streaming data. However, the streaming data is usually too large to process as raw data in BSN environment because the nodes in BSN have a small memory and processing capacity. Hence for the classification of patients' activities, we need the suitable time series approximation and feature extraction techniques which can be easily deployed and the patterns can be analyzed in real-time. New approximation and feature analysis algorithms will make it easier to distinguish between activities such as jogging, resting, walking and transitional activities. The results from approximation and feature extraction will be feed into the machine learning based-classification algorithms.

Some approaches have been proposed to deal with the classification of sensor data with real time approximation of the series [5], but it has not been tested in the area of BSN. There has been work done in the automatic segmentation of the time series for classification [6]. The symbolic time series representation SAX [7] has been a popular approach for a wide variety of time series data. The SAX representation allows the application of string manipulation algorithms for feature extraction. One such feature extraction method is the time series bitmap [8], which has previously been applied to sensor data analysis.

In this paper, we proposed new approximation and feature extraction algorithms for BSN. We also test the accuracy of classification of the BSN-based time series data. The activities of four classes (jogging, resting, walking and transitional activities) are recorded in segments and applied with the SAX representation. The time series density map is then obtained from the symbolic representation. This map is then input to the classification algorithms. The classification algorithms are consisted of two steps: primary classification and secondary classification. The effectiveness of our proposed approach is validated using the percentage accuracy obtained from cross-validation.

2 Proposed Approach

The proposed approach in our research is divided into three steps: data collection, feature extraction, and classification. The BSN we use include two types of nodes: the sender and the receiver which communicate wirelessly with each other [9]. The sender has the accelerometer sensor attached to it, and it is controlled with the use of a sampling program available with the BSN. The sender node is fixed to the volunteer's hand and the acceleration data of the hand movement is recorded for a period of 8 seconds while he/she performs the following activities: jogging, walking, resting, picking up and putting down things, transitional activities (lying to sitting and, sitting to standing). This data is stored offline for further use. Fig. 1 shows the steps involved in this approach.

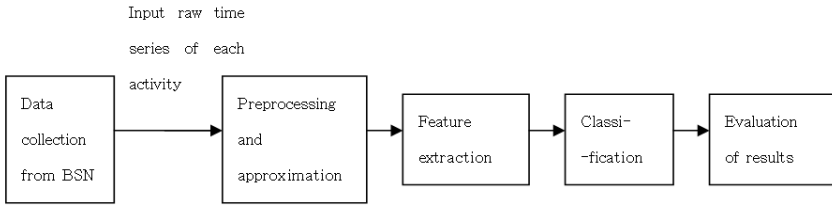


Fig. 1. The overview of our proposed approach

2.1 Data Collection from BSN

This is the first step in our proposed approach. The process of data collection involves setting up the BSN and recording of the acceleration of the volunteer's hand movements while the four classes of activity (jogging, walking, resting, and transitional) are being performed. The X, Y and Z acceleration data are recorded in a CSV file for as long as eight seconds. The yellow curve (Series 3) of Figure 2 represents the X acceleration. The pink curve (Series 2) of Figure 2 indicates the Y acceleration. The blue curve (Series 1) of Figure 2 is the Z acceleration. Thus for each class we record a collection of 100 instances. A sample time chart of the jogging activity is shown in Fig. 2.

2.2 Preprocessing and Approximation

Our second step is to adopt SAX like algorithm [7] to the time series of each activity. The Sax algorithm converts the time series into a symbolic representation. The advantages of using SAX include dimensionality reduction, discretization, lower distance bounding and numerosity reduction which makes it apt for the body sensor network data. The procedure is as follows: first the time series is normalized to remove noise, the time series of length n is divided into w equal parts and the average of each part is calculated. This is piecewise aggregate approximation (PAA) of the sequence. The cutpoints are determined from the Gaussian look up table. Usually the number of cutpoints is set as 3 and hence 4 symbols would be used for time series. Then the PAA which lies in the specified cut point slot is assigned a unique symbol or character. Thus we obtain the symbolic representation of the time series. For more details of the SAX algorithm, please refer to the original paper [7].

2.3 Feature Extraction

In the third step of our proposed approach, the SAX approximation is converted to a time series density map or bitmap. This approach is particularly useful while streaming real time data for classification. It is small and efficient enough to be handled by the BSN process or itself. The algorithm is described as follows. The Sax representation from the previous step (second step) has its alphabet size as four. We

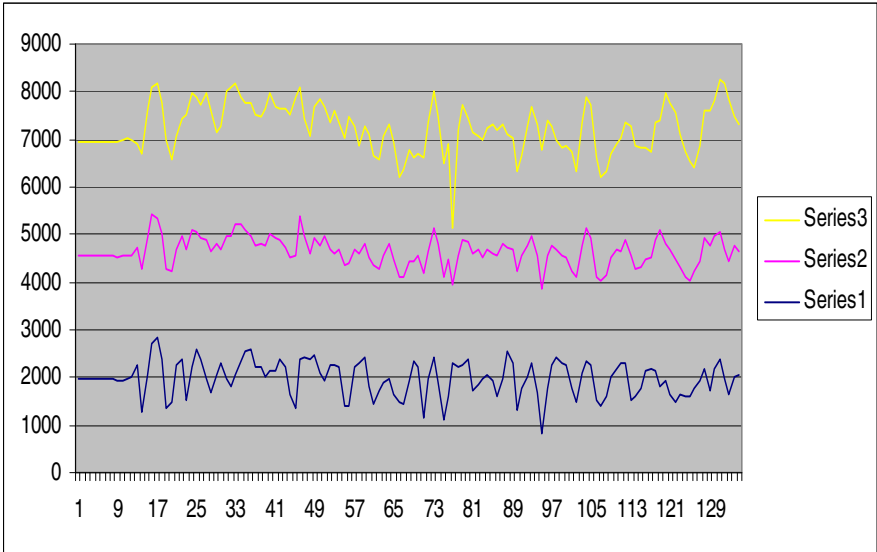


Fig. 2. The time chart for jogging activity. The yellow curve (Series 3) represents the X acceleration. The pink curve (Series 2) indicates the Y acceleration. The blue curve (Series 1) is the Z acceleration.

assign each alphabet a key. For example if A, G, C and T were the symbols used, they are assigned key values 1, 2, 3 and 4. Now the SAX representation is converted into a time series density map by moving a sliding window across the pattern and updating a frequency matrix of length $2^l \times 2^l$. The value of l is set as three since it has been found to work best in many experiments. It is also the length of the combination of symbols used in the Sax representation. An example of the matrix for length $l=1, 2$ and 3 is shown in Fig. 3.

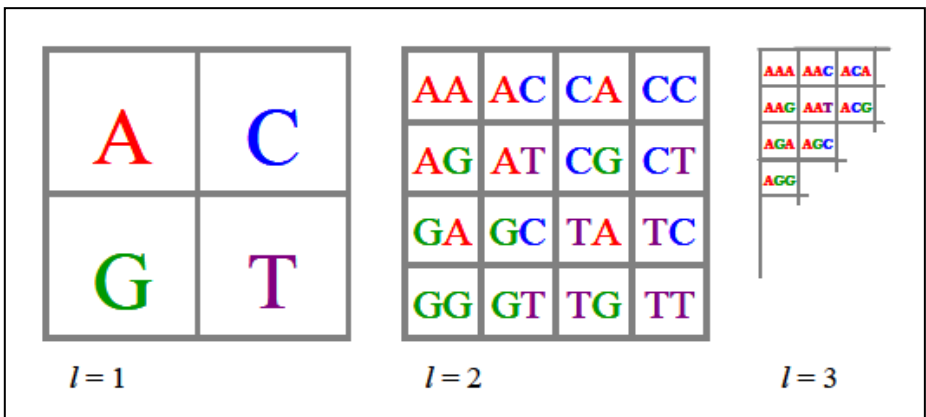


Fig. 3. The mapping of words of $l = 1, 2$ and 3

Suppose $l=4$, the first word in the sequence of the example string GATCACAGGTCTATCACCC is GATC. Let $k_0 = G$, $k_1 = A$, $k_2 = T$ and $k_3 = C$ be the key values. To map this word to a bitmap we use the formula,

$$col = \sum_{n=0}^{l-1} (k_n * 2^{l-n-1}) \bmod 2^{l-n} \quad row = \sum_{n=0}^{l-1} (k_n / 2) * 2^{l-n-1} .$$

Thus we obtain

the density map of patterns in the sequence or the bitmap of the sequence, by sliding a window of length l through it and updating the matrix based on the frequency of the possible combinations in the sequence. The output is an 8×8 matrix of cells which is time series density map of the specific activity instance. A training file is thus created with the 64 values as attributes for all the 400 instances of activity, including their class specified. This is fed as input to machine learning-based classifier.

2.4 Classification

This is the last step of our proposed approach. In order to evaluate the classification performance of the proposed approximation and feature extraction algorithms (step 2 and step 3), six machine learning-based classification algorithms were chosen. They are Bayes net, Naïve bayes multinomial, Simple logistic, Logit boost, Rotation forest, and Random forest. The classification was divided into primary and secondary steps. Primary classification performed the upper level classification for the four activities. The secondary classification performed lower division classification on the output of the primary i.e. the transitional activities were further classified into working (data pick up and put down) and transition activities (lying to sitting and sitting to stand). Training and cross validation testing are then performed on the training file. The evaluation of results is discussed next in the paper.

3 Experimental Result

3.1 Data Input to Classifier

The training data file contained 400 instances in total with four class divisions and 64 attributes. This was for the primary classification. The file for the secondary classification contained 200 instances with two class divisions and 64 attributes.

3.2 Classification Results

The classification results for the primary and secondary are shown in Table 1 and Table 2. Six classifiers have been used. The X, Y, and Z classification have been performed separately and their average result is also shown for number of folds as 8 for primary and 12 for secondary. The primary classification receives around 90 percent accuracy for all classifiers, while the secondary classification achieves only 70 percent. This shows that if the actions are well defined the time series density map would make a good choice for feature extraction in body sensor networks.

Table 1. Percentage classified correctly for primary classification

Eight Folder Cross Validation	Bayes Net	Naïve bayes Multinomial	Simple Logistic	Logit Boost	Rotation Forest	Random Forest
accelerometer X	90	90	90	88	90	90
accelerometer Y	90	90	90	90	88	89
accelerometer Z	90	88	91	89	89	89
average	90	89	90	89	89	89

Table 2. Percentage classified correctly for secondary classification

Eight Folder Cross Validation	Bayes Net	Naïve bayes Multinomial	Simple Logistic	Logit Boost	Rotation Forest	Random Forest
accelerometer X	67	71	74	71	76	75
accelerometer Y	48	63	85	78	82	77
accelerometer Z	48	57	62	58	67	68
average	54	64	74	69	75	73

4 Conclusion

We have shown in this paper that the time series density maps feature extraction method works well for the body sensor network data. It has also been estimated that this method has a high percentage of accuracy when it comes to classifying body movements and activities of BSN data. The feature extraction technique used here can be easily deployed on the BSN, enabling quick, time and space efficient and accurate activity classification with streaming sensor data. This can be implemented as future work, where we can use more sensors and nodes and improve the accuracy further by using combined data. We will also investigate the new feature extraction and classification algorithms under challenging environments, such missing sensor data, low quality of service, and unreliable sensor data. Our ultimate goal is to develop effective and efficient classifier for real-world clinical applications and thus serve a good purpose in health monitoring systems.

References

- [1] Yicka, J., et al.: Wireless sensor network survey. *Computer Networks* 5, 2292–2330 (2008)
- [2] Su, W., et al.: Wireless sensor network survey. *Wireless Sensor Networks: a Survey* 38, 393–422 (2002)
- [3] Yang, G.-Z.: *Body Sensor Networks*. Springer Science+Business Media LLC, New York (2006)
- [4] Schmidt, R., et al.: Body Area Network BAN—a key infrastructure element for patient-centered medical applications. *Biomedizinische Technik. Biomedical Engineering* 47, 365–368 (2002)

- [5] Kasetty, S., et al.: Real-Time Classification of Streaming Sensor Data. In: Proc. of 20th IEEE Int'l Conference on Tools with Artificial Intelligence (2008)
- [6] Guenterberg, E., et al.: An Automatic Segmentation Technique in Body Sensor Networks Based on Signal Energy. In: Proc. of The Fourth International Conference on Body Area Networks (BodyNets) (2009)
- [7] Lin, J., et al.: A symbolic representation of time series, with implications for streaming algorithms. In: Proc. of the 8th SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, San Diego, California, USA (2003)
- [8] Kumar, N., et al.: Time-series Bitmaps: a Practical Visualization Tool for Working with Large Time Series Databases. In: Proc. of SIAM 2005 Data Mining Conference, Newport Beach, California, USA (2005)
- [9] Yang, G.-Z., et al.: Tutorial on Body Sensor Networks (2010),
<http://vip.doc.ic.ac.uk/bsn/tutorial>