Improving Activity Recognition for Missing Data

Tahera Hossain, Goto Hiroki, and Sozo Inoue

Kyushu Institute of Technology
1-1 Sensui-cho, Tobata-ku, Kitakyushu-shi, Fukuoka, 804-8550, JAPAN
{tahera|goto|sozo}@sozolab.jp

Abstract. Automatic recognition of physical activities commonly referred to as Human Activity Recognition (HAR)—has emerged as a key research area in Human-Computer Interaction and mobile and ubiquitous computing. To monitoring activities as well as to overcome network level challenges for providing good facilities in smart care system is very important in this case. Noise and missing data are intrinsic characteristics of real world data. Monitoring activities in smart care system are a difficult task. The main challenges of caregiving facilities for detect critical activity recognition is data loss/missing through various sensors. In this paper, we analyze sensor data for detecting human activities. We focus on improvement of recognition performance in the missing data environment. We create randomly data missing environment in our dataset and try to observe the performance of accurate activity recognition rate on the missing data situation. By using our algorithm for handling missing data in our experiments it shows that recognition performance increases and it becomes 71% while 10% data has missing data in the training and testing dataset by our proposed method.

Keywords: Activity recognition, Sensor network, Wearable sensors, Machine learning.

1 Introduction

In recent years, elderly population in developed countries is increasing. According to the Population Reference Bureau [1], over the next 20 years, the 65-and over population in the developed countries will become almost 20% of the total population. Hence, the need to provide quality care and service. Now in this situation the continuous assessment of physical activities can provide an early indication of decline in health by using wearable sensors-based system.

Wearable sensors can be extremely useful in providing accurate and reliable information on people's activities and behaviors, thereby ensuring a safe and sound living environment [2]. One of the inevitable challenges of real-world data analysis is uncertainty rising from noise and missing data [3].We have to consider this missing data environment when we try to detect human activity by using sensor network. In this paper, we try to evaluate activity recognition by simulated data and observe the

performance of simulated data with and without missing data. After that we develop our algorithm for real sensor data considering a part of data as a training data and a part of data for testing data. We have used two machine learning approaches (Naïve Bayes and Random Forest) to build a model and train the dataset by calculating feature values with a particular time stamp data, mean and variance of the accelerometer data. Then we (1) generate random missing data in our test dataset, and (2) generate missing data in both train and test dataset. Afterwards, we observe the performance of recognition results with and without missing data environment. In this experiment, we replace missing data as NA in our dataset randomly. In our proposed method we consider both train and test dataset has missing data. In this work, we evaluate the performance of activity recognition detection rate assuming real life network environment constraints with data loss.

The paper is organized as follows: after introducing the works in Section 1, we present a short background in Section 2. In Section 3, we present the proposed method and methodology of our work. We explain the experimental database and results in Section 4. Section 5 concludes the paper.

2 Method

In our method, we observe that when train and test both dataset has missing data that means we train with missing data and buildup model with missing data and testing with some random missing data then it improve performance. We compare the performance result with only missing data in test dataset and missing data in both test and train dataset. If train with clean data and data is missing only in test data then activity recognition performance decrease. In our method, we find that if we train with missing data, and then test with random missing data, then our recognition performance improves compare to previous one.

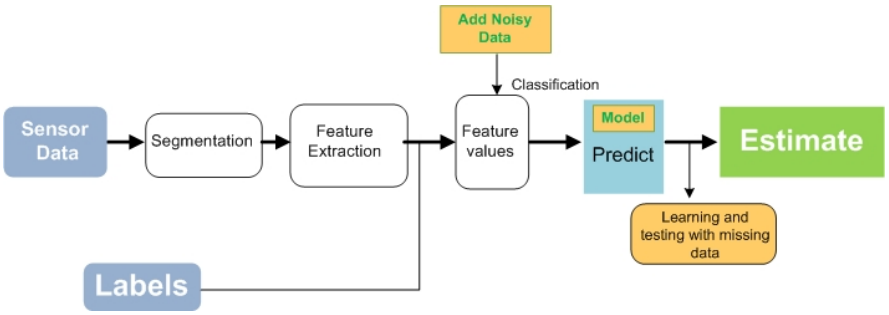

Fig. 1 Data processing method with missing data

In this paper, we have used machine learning approach to train and test the dataset by using two different classifiers called Random Forest and Naïve Bayes and found the similar experience of activity recognition performance in data missing environment.

3 Method and Experimental Setup

We have used Human Activity Sensing Consortium (HASC) challenge dataset. The dataset has more than 6700 accelerometer data with 540 subjects have been collected to develop an activity sensor dataset [4]. It has six activities, namely – stay, walk, jog, skip, stair-up, and stair-down. To collect these data, Apple iPod touch 3G is used. In our experiment, first we try to find out the activity recognition performance with simulation data in both clean and noisy environment (Table 1).

Table 1. Recognition rate by simulated data

Clean Simulation data recognition rate	Missing Data Percentage	Conventional method (%)	Proposed method (%)
100%	20%	83.33	86.66
	40%	71.7	85
	50	55	83.3
	70	38.3	50

Then we use HASC dataset for testing with real sensor activity data environment. By using cross-validation, we have differentiated test and training data from this dataset. Later, we have considered sequence data as test data. First, we have calculated features by considering mean and variance of the x, y, and z axes of accelerometer data. For classification, we have used naïve Bayes as classifier, as well as, Random Forest classifier. In this experiment, we use specific window time to calculate the mean and variance of data for creating features. In this work, we have used two classifiers for recognition. Table 2 provides a comparative result for recognition by using these two classifiers with clean dataset.

Table 2. Recognition results in percentage

Action Names	Naïve Bayes	Random Forest
Stay	100%	100%
Walk	93%	63%
Jog	94%	85%
Skip	96%	92%
stUp	100%	82%
stDown	0%	17%
Total Recognition Rate	85%	76%

Afterwards, we calculate recognition rate in data missing environment first only in test data with data missing environment. Then both test and train with missing data environment. We found recognition result by random forest and Naïve Bayes (Table 3) with randomly missing data environment.

Table 3. Recognition result with missing data

Missing Data Percentage	Random Forest with training: clean data, testing: missing data	Random Forest: both training and testing have missing data	Naïve Bayes: with training: clean data, testing: missing data	Naïve Bayes: both training and testing have missing data
5%	69%	70%	81%	82%
7%	70%	72%	81%	82%
8%	67%	70%	79%	80%
10%	67%	71%	78%	78%
20%	61%	61%	72%	73%

4 Conclusion and Future work

In this paper, we try to evaluate activity performance result with noisy missing data environment with two classifiers. By simulated data recognition accuracy was improved by our proposed method from 83.33% to 86.66% when we test with 20% missing data. In our experiment, we observe that accuracy of recognition rate improve to 71% with 10% missing data by our proposed method. While previous time it was 67% with 10% missing data. This 10% missing data has activity labels but data values have no data for data missing. In our result analysis, we found that our classifier could not recognize most of the stDown activities. We will focus this part in future. We have calculated mean and variance of data for calculating feature value. We think that other statistical measure will give some more accurate result for sensor based activity recognition which we will use in future. In future, we will try to improve accuracy result more than present.

References

1. K. Kinsella, and D.R. Phillips, “Global aging: The challenge of success”, Pop. Bull, vol. 60, 1-42, 2005.

2. J. Yin, Q. Yang, and J. Junfeng Pan, “Sensor-Based Abnormal Human-Activity Detection”, IEEE Transactions on Knowledge and Data Engineering, vol. 20, no. 8, 2008.

3. Z. Ghahramani. 2015. Probabilistic Machine Learning and Artificial Intelligence. Nature 521, 7553 (2015), 452–459.

4. N. Kawaguchi, N. Ogawa, Y. Iwasaki, K. Kaji, T. Terada, K. Murao, S. Inoue, Y., Kawahara, Y. Sumi, and N. Nishio, “HASC Challenge: Gathering Large Scale Human Activity Corpus for the Real-World Activity Understandings”, Pr of ACM AH 2011, pp. 27:1-27:5, 2011.