A Recurrent Neural Network with Non-gesture Rejection Model for Recognizing Gestures with Smartphone Sensors

Myeong-Chun Lee and Sung-Bae Cho

Dept. of Computer Science, Yonsei University
50 Yonsei-ro, Seodaemun-gu, Seoul 120-749, Korea
lmspring@sclab.yonsei.ac.kr, sbcho@cs.yonsei.ac.kr

Abstract. Gesture recognition provides a new interface to user. Various methods for the gesture recognition are feasible in smartphone environment since a number of sensors attached are gradually increasing. In this paper, we propose a gesture recognition method using smartphone accelerometer sensors. The high false-positive rate is definite if the gesture sequence data are increased. We have modified BLSTM (Bidirectional Long Short-Term Memory) recurrent neural network with non-gesture rejection model to deal with the problem. A BLSTM model classifies the input into the gesture and non-gesture classes, and the specific BLSTM models for the gestures further classify it into one of twenty gestures. 24,850 sequence data are used for the experiment, and it consists of 11,885 gesture sequences and 12,965 non-gesture sequences. The proposed method shows higher accuracy than the standard BLSTM.

Keywords: Smartphone Accelerometer Sensors, Gesture Spotting, Gesture Recognition, Recurrent Neural Network.

1 Introduction

Recently, various sensors such as accelerometer, ambient light, proximity, dual cameras, GPS, dual microphones, compass, and gyroscope are attached in a smart phone. Those not only are sophisticated but also show good performance. It is possible to recognize the user's pattern or aware the user's intention from the smartphone sensors. Especially, accelerometer is one of the most commonly used sensors for the physical movements of the user bringing the phone together. For this reason, much user interface works such as gesture recognition, and activity recognition are proposed [1].

However, there are two crucial problems when user recognition systems are developed from the smartphone sensors. One is the recognition of the non-gesture or non-activity data. The data for gesture or activity consist of meaningful and non-meaningful parts. Non-meaningful data can be half of total data or even more than gesture data. In this case, it is time-consuming to recognize both meaningful and non-meaningful data. The other is to maintain a high-accuracy rate even though the number of classes is increased. Most works about pattern recognition use machine learning methods to learn

the data. However, the performance degrades when classifying a large number of classes. It is the important problem when supporting the various services to users.

In this paper, we propose a mobile gesture recognition method where data is collected from an accelerometer sensor attached to smartphone. To alleviate aforementioned two problems, a recurrent neural network called BLSTM (Bidirectional Long Short-Term Memory) is used for classifying twenty gestures. To distinguish the gesture and non-gesture, a non-gesture rejection model is proposed.

2 Related Works

The relevant research about gesture recognition is divided into several groups. One is the vision-based gesture recognition. Q. Chen et al. proposed a hand gesture recognition method with the combination of statistical and syntactic analyses. They used a web camera as the input device and constructed the system for real time environment [2]. Tran et al. suggested a 3-D posture tracking for real time hand gestures. They used nearest neighbor clustering algorithm based on LCS (Longest Common Subsequence) similarity measure of joint angle dynamics [3]. Next group of methods is to recognize gestures based on EMG (Electromyogram) sensor, which is a device to measure the electrical potentials generated by muscle cells. Zhang et al. used both accelerometer and EMG to classify 72 Chinese Sign Language (CSL) [4]. Wheeler et al. utilized the EMG to recognize their arm movement to control a joystick. The Bayesian decomposition method is proposed to distinguish individual muscle groups with the goal of enhancing gesture recognition [5].

Another method recognizes gestures based on accelerometer. It is divided into two groups again, using accelerometer chip only and using the accelerometer attached in a smartphone. Xu et al. used MEMS 3-axes accelerometer for seven hand gestures with the Hopfield network as associative memory for classification algorithm. A decision tree and multi-stream HMMs (Hidden Markov Models) were utilized as decision-level fusion to get classification result [6]. Each method has merits and demerits. In case of the vision-based approach, it is possible to recognize the multiple objects. However, it has a limitation to record data because of the constraints of the camera placement. The detailed data can be collected from EMG but the data cannot be collected wirelessly. Accelerometer is more suitable device to collect the gesture data since it is used without the constraints of place and can be collected the gesture data accurately. For our problem, we use an accelerometer in a smartphone. The related works using accelerometer are shown in Table 1.

Table 1. Related Workd using Mobile Accelerometer

Author	Year	Classifier	Gestures	Collector	Overview
Liu <i>et al.</i> [7]	2009	DTW	8	Smartphone	Personalized gesture using uWave
Niezen <i>et al.</i> [8]	2009	HMM, ANN, DTW	8	Smartphone	Performance comparison between many classifiers

Table 1. (Continued)

Author	Year	Classifier	Gestures	Collector	Overview
Min <i>et al.</i> [9]	2010	DTW, NB, K-means	20	Smartphone	DTW model selection through NB
Marasovic <i>et al.</i> [10]	2011	K-NN	7	Smartphone	Combination of the PCA and K-NN
Akl <i>et al.</i> [11]	2011	DTW, AP	18	Wii mote	Dimensionality reduction through RP

Fig. 1. Overview of the proposed method

3 The Proposed Method

This paper aims to enhance the accuracy by using non-gesture rejection model. The entire system configuration is shown in Fig 1. The accelerometer data collected from a smart phone are segmented by using sliding window and average variation. The preprocessed data are classified after training. We adopt the recurrent neural network based on BLSTM which is a hybridization of BRNN (bidirectional recurrent neural network) and LSTM. A set of training data is used to classify the gestures and non-gestures, and the other set of training data for classifying the twenty gesture classes.

LSTM is an extension of the recurrent neural network. It uses the three gates that can store and access the data collected from the rest of the network. The gates are activated from logistic sigmoid activation function. The performance outweighs HMM used for time-series classification problem. The Hyperbolic tangent activation function is used for squashing functions. The basic calculation of each gate is the same with standard artificial neural network [12].

The input gate determines whether the input values put the memory cell or not.

$$\alpha_i^t = \sum_{i=1}^I w_{ii} x_i^t + \sum_{h=1}^H w_{hi} \beta_h^{t-1} + \sum_{c=1}^C w_{ci} s_c^{t-1} \quad (1)$$

$$\beta_i^t = f(\alpha_i^t)$$

where α_i^t is a state of the input gate at time t. It is calculated from input values, the output of other networks, and state of the memory cell. I, H, and C mean the numbers of input node, hidden node, and cell, respectively. w is the weight of connected nodes. $f(x)$ is the logistic sigmoid function to activate the input gate.

The output gate determines whether the information is output or not. The calculation of the output gate is similar with the input gate.

$$\alpha_\phi^t = \sum_{i=1}^I w_{i\phi} x_i^t + \sum_{h=1}^H w_{h\phi} \beta_h^{t-1} + \sum_{c=1}^C w_{c\phi} s_c^{t-1} \quad (2)$$

$$\beta_\phi^t = f(\alpha_\phi^t)$$

where α_ϕ^t is a state of the forget gate at time t and β_ϕ^t is a state after applying the activation function.

Eq. (3) is a calculation that is generated by forget gate, the state of cell, the state of input gate, and state of α_c^t after applying hyperbolic tangent activation function. Note that the fixed weight value 1.0 is used for preserving the information in the memory cell.

$$\alpha_c^t = \sum_{i=1}^I w_{ic} x_i^t + \sum_{h=1}^H w_{hc} \beta_h^{t-1} \quad (3)$$

$$s_c^t = \beta_\phi^t s_c^{t-1} + \beta_i^t g(\alpha_c^t)$$

Forget gate provides the information to reset the memory cell.

$$\alpha_\omega^t = \sum_{i=1}^I w_{i\omega} \alpha x_i^t + \sum_{h=1}^H w_{h\omega} \beta_h^{t-1} + \sum_{c=1}^C w_{c\omega} s_c^{t-1} \quad (4)$$

$$\beta_\omega^t = f(\alpha_\omega^t)$$

where ω is the output gate and β_{ω}^t is the state of an output gate after applying the activation function at time t .

Eq. (5) is the definition of the cell output. To activate the cells, hyperbolic tangent is used and multiplies the state of output gate.

$$\beta_c^t = \beta_{\omega}^t e(s_c^t) \quad (5)$$

For training the LSTM recurrent neural network, we use the Back Propagation Through Time (BPTT) algorithm.

4 Experimental Results

For the experiments, Samsung Omnia smartphone is used with MS Windows Mobile 6.1 as a platform. The acceleration is sampled at 50Hz. 30 people of 10~60 years old participate in the experiments. The collected data are divided into generations and date. Total amount of the data consists of 11,885 gesture sequences and 12,965 non-gesture sequences. The number of files which is used in the experiment is 1,075.

Table 2 shows the symbol of twenty gestures, direction and meaning. Rotating and tilting to hold their physical states after the movement. Tapping represents the hand or finger stroke on a smart phone surface. In the case of shaking, subjects shake the devices two or more times in a specific direction. Snapping has an angular acceleration while bouncing moves straightly to a direction and reflected back where both are the kind of pendulum movement.

Table 2. Types of Gesture Classes

Symbol	Direction	Meaning	Symbol	Direction	Meaning
NL	Left	Snapping	TL	Left	Tapping
NR	Right		TR	Right	
NF	Forward		TF	Forward	
NB	Backward		TB	Backward	
BU	Up	Bouncing	TT	Top	Rotating
BD	Down		TM	Bottom	
LL	Left	Tilting	RH	Horizontal	
LR	Right		RV	Vertical	Shaking
LF	Forward		SLR	Left-Right	
LB	Backward		SFB	Forward-Backward	

For the first experiment, the data are divided into a ratio of seven to three for training and test, respectively. The 17,470 sequences are used for training, and 7,380 sequences are used for test. Each sequence is distributed randomly. The results through all experiments in this work are compared with standard BLSTM. Average accuracy rate of the proposed method is 91.11% and standard BLSTM is 89.17%. For the second experiment, we group the data as generation. 18,490 sequences are used for training and approximately 2,100 sequences are used for testing each generation. Fig 2 shows that the proposed method outperforms standard BLSTM for all generations.

Fig. 2. The generational experiment

5 Concluding Remarks

In this paper, we have collected the accelerometer data from a smartphone and classified the data by using a recurrent neural network with non-gesture rejection model. Total classes of the data are twenty one including non-gestures. The performance of the standard BLSTM compared with the proposed method and our approach outperforms the standard BLSTM. However, it is not satisfactory to apply in the real world. It might be possible to obtain higher accuracy if the data are grouped with the similar meaning because some gestures have similar features. For a future work, we will investigate the characteristics of features for similar gestures.

References

1. Lane, N.D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., Cambell, A.T.: A survey of mobile phone sensing. *IEEE Communications Magazine* 48(9), 140–150 (2010)
2. Chen, Q., Georganas, N.D., Petriu, E.M.: Hand gesture recognition using haar-like features and a stochastic context-free grammar. *IEEE Trans. on Instrumentation and Measurement* 57(8), 1562–1571 (2008)
3. Tran, C., Trivedi, M.M.: 3-D posture and gesture recognition for interactivity in smart spaces. *IEEE Trans. on Industrial Informatics* 8(1), 178–187 (2012)
4. Zhang, X., Chen, X., Li, Y., Lantz, V., Wang, K., Yang, J.: A framework for hand gesture recognition based on accelerometer and EMG sensors. *IEEE Trans. on Systems, Man, and Cybernetics, Part A: Systems and Humans* 41(6), 1064–1076 (2011)
5. Wheeler, K.R., Chang, M.H., Knuth, K.H.: Gesture-based control and EMG decomposition. *IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews* 36(4), 503–514 (2006)
6. Xu, R., Zhou, S., Li, W.J.: MEMS accelerometer based nonspecific-user hand gesture recognition. *IEEE Sensors Journal* 12(5), 1166–1173 (2012)
7. Liu, J., Wang, Z., Zhong, L., Wickramasuriya, J., Vasudevan, V.: uWave: accelerometer-based personalized gesture recognition and its applications. *IEEE Int. Conf. on Pervasive Computing and Communications*, 1–9 (2009)

8. Niezen, G., Hancke, G.P.: Evaluating and optimising accelerometer-based gesture recognition techniques for mobile devices. *AFRICON*, 1–6 (2009)
9. Min, J.-K., Choe, B.-W., Cho, S.-B.: A selective template matching algorithm for short and intuitive gesture UI of accelerometer-builtin mobile phones. In: *Cong. on Nature and Biologically Inspired Computing*, pp. 660–665 (2010)
10. Marasovic, T., Papic, V.: Accelerometer-Based Gesture Classification Using Principal Component Analysis. In: *Int. Conf. on Software, Telecommunications and Computer Networks*, pp. 1–5 (2011)
11. Akl, A., Feng, C., Valaee, S.: A novel accelerometer-based gesture recognition system. *IEEE Trans. on Signal Processing* 59(12), 6197–6205 (2011)
12. Hochreiter, S., Schmidhuer, J.: Long short-term memory. *Neural Computation* 9(8), 1735–1780 (1997)