Gamification for High-Quality Dataset in Mobile Activity Recognition

Nattaya Mairittha(✉) and Sozo Inoue

Kyushu Institute of Technology, 1-1 Sensui-cho, Tobata-ku, Kitakyushu-shi, Fukuoka 804-8550, Japan
{fah,sozo}@sozolab.jp

Abstract. This paper presents a gamification concept for getting high-quality user-annotated datasets in the context of mobile activity recognition, as well as a cheating detection algorithm. The novel idea behind this concept is that users are motivated by getting feedback about the quality of their labeling activity as rewards or gamification element. For that, the collected sensor data and labels are used as training data for a machine learning algorithm for determining the dataset quality based on the resulting accuracy. By using the proposed method, the results show that the gamification elements increase the quantity (labels from the proposed method is higher than the naive by at least 305) and the quality (the accuracy of the proposed data outperformed the original data by at least 4.3%) of the labels. Besides, the cheating detection algorithm could detect cheating with the accuracy of more than 70% that is fascinating work.

Keywords: Mobile activity recognition · Quality of dataset Gamification

1 Introduction

Mobile activity recognition is the technology of recognizing human activities with mobile sensors such as smartphones. It is widely researched [1], as in the preventive healthcare domain and process management and skill assessment of workers. To address the mobile activity recognition task [2], collecting high-quality training datasets with correct ground truth labels is very costly and non-trivial task. In many real-world situations, the number of training examples must be limited because obtaining samples in a form suitable for learning may be costly [3]. These costs include the cost of collecting the raw data, cleaning, storing and transforming the data into a representation suitable for learning, as well as the opportunity cost associated with suboptimal learning from large datasets due to limited computational resources [4].

To collect good quality of labels without tedious/costly tasks and limit disengagement., L'Heureux [5] proposed to motivate users to participate in various labeling or tasks. As such, increasing the stimulus of the labeling task itself

K. Murao et al. (Eds.): MobiCASE 2018, LNICST 240, pp. 216–222, 2018.
https://doi.org/10.1007/978-3-319-90740-6_14

can address tedious tasks caused by user disengagement. Studies by Markey [6] revealed four effective strategies associated with heightened task engagement: offer performance feedback, provide social approval, increase challenge, and give incentives such as monetary rewards. Gamification, Which is commonly defined as *the use of game design elements in non-game contexts to improve user experience* [7], puts these strategies to advantage. Research on gamification has shown improvements in motivation and engagement [8]. Game elements such as progress and success feedback, goals, points, badges, levels, challenges, social feedback, and narrative, can all contribute to those engagement improvements, provided there is a good match between design and audience. Therefore, gamification offers a useful perspective from which to create and analyze engaging labeling experiences. Here, we can come up with the idea of if we synchronize the goal of gamification and the quality of the dataset. Then, we can expect that we could motivate people to provide high a quality dataset for activity recognition.

The contribution of our work is twofold. First is to get the high-quality datasets in the context of mobile activity recognition by exploiting gamification concept. We present an idea that users provide sensor data and activity labels as a training dataset, as well as obtaining gaming feedback as a gaming element. However, giving material rewards as motivators to drive specific user actions. One of the side effects of users getting too focused on the rewards because these motivators are tangible, visible and highly desirable [9]. For example, users who are more interested in rewards than in physical activity might cheat by labeling the data walk by without actually stepping it or labeling run by sitting still at their desk. Therefore as a second contribution, to prevent cheating, we propose to defeat the cheats by making algorithms detecting cheating, based on the assumption that cheating datasets are dissimilar to other (non-cheating) datasets.

As a result of evaluating our prototype system that provides an estimated labeling quality to users as notifications every 30 min with ten volunteers, the number of labels for the proposed method was greater than the naive by at least 305. Moreover, the quality of labels by the accuracy of the proposed data outperformed the original data by at least 4.3%. In addition, the proposed method detected cheating data with the accuracy of more than 70%.

2 Proposed Gamification System

2.1 Gamification Mechanism

The first goal of the integration is to synchronize the goal of gamification and the quality of the dataset by providing the 'point' (or a score) as a reward for each the user, which is the one that each user wants to maximize. On the other hand, the quality of the dataset can be represented by the accuracy of the activity dataset when they are trained with several machine learning algorithms. So, the first idea is *"to let the accuracy of the dataset of a user be the score of her/him."* The second idea is how gamification addresses the engagement. By giving feedback to the users periodically as notifications (Fig. 1).

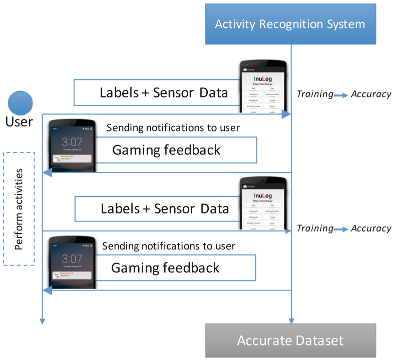

Fig. 1. The overview of proposed method

2.2 Cheating Detection

To avoid cheating by the participants, we also propose a cheating detection algorithm based on supervised machine learning. By letting all users cheat intentionally, for example, labeling the data, such as running or by sitting still at their desk and pretending to be walking. Then, we train from the dataset with intentionally cheated, and with standard gamification which we assume they are not cheating. As a preprocessing, we use the data sample with activity labels, and as post-processing, we use voting to detect 'cheating' or 'non-cheating'.

In the following, we show the training and detection algorithms.

Training Algorithm:

- Input: sensor dataset X of size N, and the same size of labels

$$C = \{\text{'cheating'}, \text{'non-cheating'}\}^N$$

.

- Output: cheating estimation function f.

1. Remove samples with no activity labels from X and C.
2. Calculate feature vectors V from X.
3. Using supervised machine learning, train a model f with X and C to estimate cheating or non-cheating.
4. Output f.

Detection Algorithm:

- Input: sensor dataset X and function f.
- Output: 'cheating' or 'non-cheating'.

1. Remove samples with no activity labels from X.
2. Calculate feature vectors V from X.
3. Using f, estimate cheating or not by $f(v_i)$ for $\forall v_i \in V$.
4. Output the maximum voting results by

$$o = \arg_{y_i} \max_i f(y_i).$$

The output o is the detection results. If it is detected as 'cheating', we can consider fewer rewards or penalty to that user in the framework of gamification.

3 Evaluation Experiment

3.1 Evaluation Experiment

In this section, we evaluate the effectiveness of the proposed method by answering the following questions:

- Can the proposed method improve the data quality?
- Can the proposed method detect when users in the system are cheating?

Experiment Design

We split the subjects into two groups: one is the proposed or gamified group by getting feedback about the quality of their labeling activity as rewards or gamification element, and the other is non-gamified. By randomly assigned a participant into either of two groups in one day and switched them into other groups after on the second day of the experiment. To evaluate cheating detection, we asked the participants to intentionally cheat on the third day.

Labeled Data Collection

The experiment was carried out with a group of ten volunteers within an age bracket of 20–40 years, then dividing the candidates into two groups, with each group including five candidates. Each person performed wearing a smartphone with an armband. From them, we captured three-axial acceleration data.

Data Processing

In the data processing, we introduced the preprocessing stage to synchronize the times, remove artifacts, and prepare the acquired signals for feature extraction. Later, features that capture the activity characteristics are extracted from the signals within each segment. From the 3-axial acceleration data, we extracted feature vectors in the following way: At first, we divided the samples by every minute and calculated the median and standard deviation of each axis.

3.2 Evaluation Method

Data Quality

We evaluated the quality of the obtained data using supervised machine learning and by seeing several accuracy measures. For the machine learning algorithms, we used Random forests are an ensemble learning method for classification which is popular in achieving reasonable performance. Importantly, in evaluating accuracies, to take care of imbalances among activity classes, we adopt two countermeasures: first is to use one-class classification: to classify a specific activity class or not and repeat it for any activity class, and the second is to use also imbalance-robust metrics such as *Balanced Classification Rate (BCR)* [10]. The BCR is defined as follows:

$$\mathrm{BCR} = \frac{\text{TP-rate} + \text{TN-rate}}{2}$$

Cheating Detection

To evaluate the cheating detection carefully, we exploited the cross-validation in a novel way. The details are described below:

1. From users U, take a pair of users (u_1, u_2) where $u_1 \neq u_2$,
2. take 'cheating' data from user u_1, and take 'non-cheating' data from user u_2, and let the merged data the test dataset D_E.
3. From the rest of the users $U - \{u_1, u_2\}$, take 'cheating' and 'non-cheating' data from D, and let them the training dataset D_T.
4. Train a with D_T by the cheating algorithm.
5. Estimate and take the maximum voting with D_E the cheating algorithm.
6. Repeat 1–5 to any pairs of users and sum up the results.

4 Results

4.1 Overview of the Obtained Data

Table 1 illustrates the number of labels per activity of the experiment.

Table 1. The number of labels per activity

No.	Activity class	#labels	No.	Activity class	#labels	No.	Activity class	#labels
1	Sleeping	22	8	Walking	112	15	Standing	39
2	Watching	10	9	Riding elevator	34	16	Carrying	20
3	Working on computer	54	10	Eating	26	17	Drinking	28
4	Reading	12	11	Cycling	18	18	Relaxing	20
5	Climbing stairs	25	12	Ridding escalator	12	19	Taking a bus	279
6	Taking a train	8	13	Sitting	50	20	Use the toilet	28
7	Washing	24	14	Dressing	6	21	Uses the phone	11
						22	Meeting	15

4.2 Improvement of Data Quality

Figure 2 shows the number of labels in each of the groups in our study, split per condition. As the results show that there are more labels in the gamified condition than other conditions. the number of labels for the proposed method from Group A is greater than the naive by 305, and the number of labels for the proposed method from Group B is greater than the naive by 109.

Figure 3 shows the results of accuracies for naive and proposed method. The BCR of the proposed method is greater than the naive by 4.32% the f-measure of the proposed method is greater than the naive by 27% the precision of the proposed method is greater than the naive by 26.6% the recall of the proposed method is greater than the naive by 8.02%

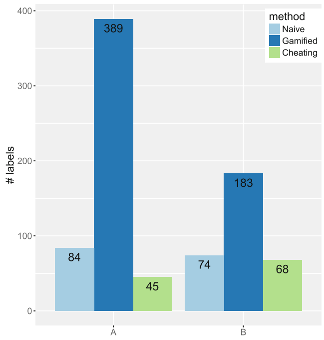

Fig. 2. Number of labels per condition

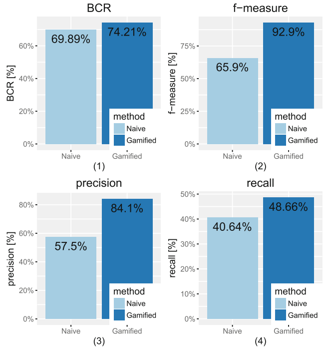

Fig. 3. Accuracies of methods

4.3 Cheating Detection

Table 2 is the confusion matrix of cheating detection after 1-pair-of-user-left-out cross-validation described in Sect. 3.2. From the table, we can calculate that the accuracy is 74.4%, the precision is 70.0%, the recall is 85.7%, the f-measure is 77.1%, and the BCR is 74.4%.

Table 2. Confusion matrix of cheating detection.

Truth\estimate	Non-cheating	Cheating
Non-cheating	31	18
Cheating	7	42

5 Conclusion

This paper aims to get high-quality user-annotated datasets in the context of mobile activity recognition by exploiting gamification concept. The results show that the gamification elements increase the quantity and ‘quality’ of the labels. Besides, the cheating detecting algorithm is fascinating work; we could detect cheating with the accuracy of more than 70%. Future research includes analyzing differences between users, groups, classes, as well as future research, should target a large in-field study such as applying crowdsourcing.

References

1. Kwapisz, J.R., Weiss, G.M., Moore, S.A.: Activity recognition using cell phone accelerometers. ACM SigKDD Explor. Newsl. **12**(2), 74–82 (2011)
2. Bao, L., Intille, S.S.: Activity recognition from user-annotated acceleration data. In: Ferscha, A., Mattern, F. (eds.) Pervasive 2004. LNCS, vol. 3001, pp. 1–17. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-24646-6_1

3. Weiss, G.M., Provost, F.: Learning when training data are costly: the effect of class distribution on tree induction. J. Artif. Intell. Res. **19**, 315–354 (2003)
4. Inoue, S., Ueda, N., Nohara, Y., Nakashima, N.: Mobile activity recognition for a whole day: recognizing real nursing activities with big dataset. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 1269–1280. ACM, September 2015
5. L'Heureux, A., Grolinger, K., Higashino, W.A., Capretz, M.A.: A gamification framework for sensor data analytics. In: 2017 IEEE International Congress on Internet of Things (ICIOT), pp. 74–81. IEEE, June 2017
6. Markey, A.R.: Three essays on boredom (2014)
7. Deterding, S., Dixon, D., Khaled, R., Nacke, L.: From game design elements to gamefulness: defining gamification. In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, pp. 9–15. ACM, September 2011
8. Hamari, J., Koivisto, J., Sarsa, H.: Does gamification work?-a literature review of empirical studies on gamification. In: 2014 47th Hawaii International Conference on System Sciences (HICSS), pp. 3025–3034. IEEE, January 2014
9. Glas, R.: Breaking reality: exploring pervasive cheating in Foursquare (2015)
10. Fawcett, T.: An introduction to ROC analysis. Pattern Recogn. Lett. **27**(8), 861–874 (2006)