Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 00 (2011) 000–000

Procedia Engineering

www.elsevier.com/locate/procedia

APCST 2011: 5th Asia-Pacific Congress on Sports Technology

Activity Classification With Smart Phones For Sports Activities

K. Taylor[ab*], U.A. Abdulla[a], R.J.N. Helmer[c], J.Lee[ab], I. Blanchonette[c]

[a]*CSIRO ICT Centre, Canberra, Australia*
[b]*Australian National University, Canberra, Australia*
[c]*CSIRO Materials Science and Engineering, Belmont, Victoria, Australia*

Abstract

Activity classification using mobile phones is useful for identifying training activities, then capturing short periods of high frequency training data and capturing and archiving appropriate training statistics for various training activities. Some available smart phone training information systems classify the negative case of resting during a training session but none actively detect training activities and classify type of activity. It is widely perceived that activity classification is useful but few activity classifiers are available for smart phones. We test one activity classifier for the Android platform that is able to run as a background application without an obvious impact on battery life and which reported high levels of accuracy. The reported accuracy was not achieved during testing, in part because users were applying different criteria to determine accuracy than developers. A smart phone classifier was developed adding several techniques to increase usefulness and accuracy as perceived by users. These included detecting device states where inferring user activity was not possible, limiting the range of activities to those that can be reliably detected, eliminating dependence on device orientation, presenting aggregated information graphically and web based archiving of activity history. The classifier can be used for detecting levels of exercise undertaken, detecting occurrence of training activities and for messaging other applications to trigger collection of appropriate detailed information and summary statistics. Combining activity information with applications inferring lifestyle activities from location based data would enhance the usefulness of both applications.

"Keywords: Activity Classification; smart phones; sports training"

[a]* Corresponding author. Tel.: +61 2 6216 7151
E-mail address: ken.taylor@csiro.au.

1. Introduction

Information systems for sports training that are smart phone based are appealing due to device familiarity, convenience and configurability. Convenience is maximised by transferring live training data to a server, so that it can be available to coaches during training. Examples of popular applications that provide some of this functionality are Runkeeper for the Android and Apple iPhone and MyTracks for the Android. MyTracks was used to stream live power, heart rate and position data during the TourDeFrance bicycle tour. Besides advantages, mobile phones impose limitations. Limitations include limited processing power, limited battery life, variable and expensive communications bandwidth and a requirement to still operate as a phone.

There is a wide range of uses for training data and different levels of detail are required for each use. For kayaking, an athlete and coach need stroke rate speed and distance data during training, biomechanists need small samples of highly detailed data to analyse style and all need summary data to monitor training levels and trends. In flat water kayaking competition and training at the Australian Institute of Sport the Minimax system running in its simplest configuration records three axis accelerometer data at 100Hz for a period that includes training preparation and all training including rest periods until completion. This data is transferred to a PC for post training processing and summarisation.

Where this model is reproduced on a mobile phone-based system communicating with a server, considering the limitations as explained before, an ideal solution is an activity classifier that can determine when training is being undertaken, record training summaries and sample portions of training sessions at high levels of detail for archiving. Where an athlete undertakes a variety of training modes all modes that can be detected can be automatically categorised and recorded.

There is also a wide range of other uses for an effective activity classification application. Google Latitude provides a good example where many aspects of a person’s life can be inferred from historical location data including where the person works and lives and how much time they spend at each location. The effectiveness of these applications would be increased by adding knowledge of a person’s activity. The health of non athletes can also benefit by measuring the amount and types of exercise they are undertaking and comparing actual activity to ideal activity levels for good health.

The smart phone applications that currently exist for sports training must be started at the commencement of a training session and stopped at completion and record the same data throughout a training session. This is required so that the battery hungry GPS functionality is only used when required and training can be separated from non training. Some applications including My Tracks and Cardio Trainer detect rest periods by the location remaining constant for a period and use this knowledge to exclude rest periods from some training statistics. This is a simple form of activity classification where the user has said, “I am training now” and the application has classified the negative case during a training session when training is not occurring. The more difficult but more useful approach is to detect when particular training activities are occurring and record and summarise data appropriately.

With the obvious advantage of activity classification it could be expected smart phone activity classifiers would be widely available but there are few. We investigated the effectiveness of an existing activity classifier for the Android smart phone identifying some limitations. We then developed an Activity Classifier based on the existing design with variations including presenting aggregated information graphically and web based archiving of activity history and report on the results.

2. Algorithms Used For Classification

Ermes *et al*. [1] using a combination of various sensors including accelerometers and GPS were able to recognize when certain sports activities were being carried out. They claimed accuracies ranging from

70% to 97%, for various sports activities. Ravi *et al*. [2] compared the performance of a number of base-level classifiers and meta-level classifiers in classifying user activities from accelerometer data. Among the compared base-level classifiers were the K-Nearest Neighbour (KNN) algorithm and naive Bayes classifier. Meta-level classifiers use the results of several base-level classifiers to increase accuracy. A meta-level classifier called plurality voting was found to yield the best results among both base-level classifiers and meta-level classifiers. Although the reported results were excellent, with a success rate of 99.57% in one experiment, it should be noted that the computation involved in meta-level classifying algorithms would be inappropriate for a battery powered mobile device. It should also be noted that the accelerometer was worn near the pelvic region by all subjects while they performed a number of selected activities and that false positives for unclassified activities was not investigated. This closed experiment setup does not compare to the usage by mobile phone users going about their day to day life. Currently we are using a KNN classifier as an input to a decision tree and aggregation filter.

3. Investigation

The activity classifier chosen for investigation is the Activity Recorder component of the Context-Aware API for the Android platform. It was chosen because it was in the Android market, able to run as a background application without an obvious impact on battery life and reported high levels of accuracy being able to “correctly classify activities with an accuracy of in excess of 95% (measured by holding back 1/3 of the training data) using the KNN algorithm with K = 1” [3]. This high level of reported accuracy was achieved with four simply calculated features. These are maximum and minimum acceleration and average acceleration over a six second sampling interval on two axes. Power consumption is considerably reduced by sampling for 6 seconds out of every 30 seconds at some cost to accuracy.

Three users unfamiliar with the application tested Activity Recorder and all were dissatisfied with the classification accuracy. An example covering the morning of a workday is given in Table 1.

In summary two walking events were correctly classified but everything else was incorrect or unknown, though it was successful at detecting the length of the incorrectly classified driving event.

This is a very different result from the 95% accuracy reported. Many other researchers report high levels of accuracy in activity classification. For example, Karantonis *et al*. [4] described an experiment where a waist-mounted triaxial accelerometer was used to classify among a series of human movement and reported an overall accuracy of 90.8%. Ermes *et al*. [1] obtained an accuracy of 90% while using researcher-annotated data for training and testing, 89% while using researcher-annotated data and user-annotated data for both training and testing, but got a much lower 72% when using researcher-annotated data for training and user-annotated data for testing. The difference shows the difference in methodology between researchers’ testing and users’ testing. As is frequently the case Karantonis et. al. [4] does not report on false positives when undertaking activities the classifier is not able to recognise. According to Consolvo et. al. [5], the types of errors can be categorised as:

1. Make an error in the start time.
2. Make an error in the duration.
3. Confuse an activity it was trained to infer with another it was trained to infer.
4. Confuse an activity it was not trained to infer with one it was trained to infer.
5. Failure to detect an activity it was trained to infer.
6. Failure to detect an activity it was not trained to infer.
7. Detect an activity when none occurred.

Table 1: Activities recorded one morning with commentary.

Activity Recorded	Actual Activity	Comments
Sitting Down for 2 mins	Phone on table.	Activity classification is unrelated to the actual human activity.
Unknown for <1 min	Not sure	Less than 1 minute is too short to know unless the screen is watched continuously.
Walking for 2 min	Not sure	Still walking around
Sitting down < 1 min	Not sitting down	Still walking around packing stuff etc
Unknown for <1 min	Not sure	Still walking around packing stuff etc
Unknown for <1 min	Not sure	
Standing still for 8 mins	Driving car phone in breast pocket.	The activity is wrong but it has recognised correctly the same activity occurred for 8 mins
Unknown for <1 min	Not sure	
Walking for 6 mins	Walking	Good
Various activities mostly sitting down for 2 hours 30mins	Phone on desk	Activity classification is obviously unrelated to the actual human activity though coincidentally correct some of the time.

Consistent with Consolvo *et al*. [5] observations, a small error in start time or duration was not noticed by people testing Activity Recorder as their perception of start time and duration is vague without intense observational effort. Consolvo *et al*. [5] stated that people found errors 5 and 7 frustrating and this "frustration often led to participants questioning if the device was malfunctioning". Of these two errors, researchers are mostly reporting the first but not the second so that reported results are more accurate than are perceived by users. This is because, as reported by Consolvo *et al*. [5] users perception of accuracy is based on different criteria than those reported by researchers. To achieve a high level of user acceptance classifiers should only report activities that can be determined with a high level of sensitivity and specificity. In the case of Activity Recorder low specificity activities are dancing, walking vs walking up or down stairs and travelling vs travelling by bus or car. Not reporting these would increase perceived accuracy.

Table 2. Activities recorded after applying rules to improve reliability.

Activity Recorded	Actual Activity	Comments
Walking for 8 minutes	Walking for 8 minutes.	Perfect result.
Walking 36 minutes	Walking 36 minutes	Perfect result.
Travelling total 16 minutes.	Travelling by car.	Perfect result. Travelling in suburban streets.
Travelling for 40 minutes. Unknown less than 1 minute	Travelling by car.	Nearly perfect result. Travelling on freeway.

Table 2 shows the results achieved when the following rules were applied:

- Keep the phone in your hip pocket,
- Combine walking and step climbing as walking,

- Combine travelling by car and bus as travelling,
- Don't undertake activities other than those known about.

This is better than the 95% accuracy reported by Smith [3] but the rules are too restrictive. The challenge is to achieve accuracies approaching these levels in practical usage scenarios.

4. Device States That Preclude Detection of Human Activity

Specificity and therefore credibility can be increased significantly by adding detection of device states that preclude detection of human activities. The easiest is battery charging which can be detected with 100% sensitivity and specificity. Another common phone state that precludes inference of human activity is when the phone is not carried, perhaps on the bedside table or on an office desk. An algorithm for detecting this was implemented which deemed a phone to be uncarried where no movement was detected and its orientation was unchanged over time. Movement is detected by comparing the standard deviation (SD) of a six second sample for each axis to two times the SD for each axis measured during calibration and the mean of each axis to the respective mean one minute earlier. Two SDs would be more than necessary but the calibration SD is below the quantization limit for the smart phones tested. For example the Google Nexus One and Sony Ericcson Xperia X10A both have a minimum resolution of 0.041m/s^2 and a SD measured during calibration is of the order of 0.03 m/s^2. This causes the sample range to be greater than would be expected for a normal noise distribution. These two conditions take precedence over the KNN classifier output in the decision tree. As phones are frequently not carried these changes remove more than half of the false positives for most people.

5. Phone Orientation

Most often when using accelerometers for measuring human activity, sensors are fixed in a specific orientation to a specific body location. Mobile phones however, are carried and used in a variety of ways and orientations which complicates activity classification. Common positions for carrying phones are breast pocket, hip pocket, jacket and handbag. Ichikawa *et al*. [6] conducted research where 419 people were interviewed in Helsinki, Milan and New York. Among the questions dealt with was how they carried their mobile phones. The result showed that for various reasons, males had a strong tendency to keep their phones in trouser pockets while females had a strong tendency to keep them in their bags. Users who tested Activity Recorder carried the handset in breast pocket, hip pocket and hand bag and the variety of locations caused many incorrect classifications.

So that the phone could be carried in any orientation the direction of gravity was measured using the mean value of each axis during sampling and each sample rotated to vertical and horizontal. The algorithm for detecting the direction of gravity will fail when the phone is rotated during sampling. To detect this condition the magnitude of the mean acceleration during sampling was compared to gravity and where it varied by more than 1m/s^2 the sample was rejected.

Previously sitting and standing were differentiated based on the phone’s orientation while in the hip pocket. In addition, it mostly relied on the user sitting while travelling. However, rotating acceleration to horizontal and vertical precludes using orientation as an indication of activity. Hence after merging sitting and standing to a single class referred to as “stationary”, testing was done where travelling was seen as still being reliably detected while travelling. However, travelling was also detected about half the time when the user was not travelling. Examining the KNN training data showed an overlap in the vertical acceleration ranges for travelling and stationary. Removing the travelling data where it overlapped with stationary removed almost all the false positives for travelling but reduced sensitivity to a few minutes in an hour of travelling.

6. Conclusion

Activity classification using mobile phones is useful for capturing short periods of high frequency training data and capturing and archiving appropriate training statistics for various training activities. Some available smart phone training information systems classify the negative case of resting during a training session but none actively detect training activities and classify type. It is widely perceived that activity classification is useful but few activity classifiers are available for smart phones. This is probably because practical mobile phone classifiers pose a more difficult problem than purpose built devices like pedometers as phones are carried in many different ways and the activity classifier can not impinge significantly on phone usability. We tested one activity classifier that is available for the Android smart phone platform and find that it is perceived to be less accurate by testers than is reported by Smith [3]. This is found to be a widespread problem with activity classifiers because users are intuitively applying different criteria to determine accuracy than developers. A smart phone classifier was developed adding several techniques to increase usefulness and accuracy as perceived by users. These included detecting device states where inferring user activity was not possible, limiting the range of activities to those that can be reliably detected, eliminating dependence on device orientation, presenting aggregated information graphically and web based archiving of activity history. The classifier could then be used for detecting levels of exercise undertaken, detecting occurrence of training activities and triggering collection of appropriate detailed information and summary statistics for detected activities. Combining activity information with applications like Google Latitude currently inferring lifestyle habits from location based data will also increase the usefulness of these applications.

References

[1] Ermes M, Parkka J, Mantyjarvi J, and Korhonen I. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *IEEE Transactions on Information Technology in Biomedicine*, 12(1):20–26, 2008.

[2] Ravi N, Dandekar N, Mysore P, and Littman ML. Activity recognition from accelerometer data. *In Proceedings of the National Conference on Artificial Intelligence, volume 20*, p. 1541. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2005.

[3] Smith C. Context-Aware API for Android Devices. PhD thesis, Imperial College London, 2010.

[4] Karantonis DM, Narayanan MR, Mathie M, Lovell NH, and Celler BG. Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. *IEEE Transactions on Information Technology in Biomedicine*, 10(1):156–167, 2006.

[5] Consolvo S, McDonald DW, Toscos T, Chen MY, Froehlich J, Harrison B, Predrag Klasnja1, *et al*. Activity Sensing in the Wild: A Field Trial of UbiFit Garden. *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*; p. 1797.

[6] Ichikawa F, Chipchase J, and Grignani R. Where's the phone? a study of mobile phone location in public spaces. In Mobile Technology, 2005, *2nd International Conference on Applications and Systems*, p. 1–8. IEEE, 2005.