Masthead Logo

The Fordham Undergraduate Research Journal

Volume 1 | Issue 1

Article 11

December 2013

Mobile Sensor Data Mining

Jeff Lockhart FCRH '13
Fordham University, furj11@fordham.edu

Mobile Sensor Data Mining

Cover Page Footnote

Jeff Lockhart, FCRH 2013, is from Phoenix, Arizona. He is a computer science and women's studies major. Jeff is currently conducting research in sensor data mining in the Fordham University computer and information science department, working with professor Gary Weiss. After graduation, Jeff plans to attend graduate school for either computer science or rhetoric.

Mobile Sensor Data Mining

by Jeff Lockhart, FCRH '13

COMPUTER SCIENCE

Introduction

At an ever increasing rate, the smartphones and other devices people carry with them in their everyday lives are packed with sensors and processing power. This provides an unprecedented opportunity to apply data mining techniques to people's activities as they go about their daily lives, without changing their routine. The goal of the Wireless Sensor Data Mining (WISDM) Project is to explore the possibilities of data mining on these powerful mobile platforms.[1] Data mining involves extracting knowledge from data using computer algorithms. A major sensor in these devices is the tri-axial accelerometer originally included for screen rotation and advanced gaming. Our work, so far, has focused on using data mining methods on the accelerometer data to identify the activities users are performing (activity recognition) while carrying the phone. Many useful applications can be built if accelerometers can be used to recognize a person's activity. We have also demonstrated that accelerometer data can be used to uniquely identify and authenticate users. While some previous work has examined sensor-based gait recognition,[2-12] our work in this communication differs in that we identify users based on the way they move during multiple activities (i.e., not just walking) using only commercially available smartphones, which are carried in the user's pocket.

In this communication, as is commonly the case, data mining is done offline by researchers who manually retrieve and prepare the data. The WISDM team is actively working to automate the process of receiving, aggregating, preprocessing, classifying, and reporting so that useful applications can be deployed to cell phone users. This automated architecture will also support future research efforts by providing a platform for data mining on mobile devices.

The WISDM project is moving ahead rapidly with over a dozen undergraduate members, in addition to a graduate student and our faculty team leader, Dr. Gary Weiss. We are continuing to submit new work to major industry conferences and broaden the project's scope. More information about the WISDM project can be found at http://www.fordham.edu/wisdm.

Thanks to Dr. Gary M. Weiss, my faculty mentor, for guiding the work and reviewing the results. Thanks also to Jennifer Kwapisz, FCRH '10, and Sam Moore, FCRH '10, who laid the groundwork for this effort. This work was financially supported by a Fordham Faculty Research Grant and a Google Faculty Research Award, and by a FCRH Summer Science Research Internship. Direct all correspondence to Jeff Lockhart at jlockhart@fordham.edu.

Experimental

A. Materials and Data

Android-based cell phones (as opposed to the iPhone) were chosen for our platform because the Android operating system is free, open-source, easy to program and already becoming a dominant entry in the cell phone marketplace. Further, Android and our data mining tools (Weka[13]) use the same programming language, Java. The WISDM project employs eight types of Android phones from several manufacturers, including Google, HTC, Motorola, and Samsung. Our devices use a range of Android OS versions from 1.5 to 2.2, a representative sample of current device-dependant diversity.

Data was collected from 53 subjects while they performed a set of pre-defined activities under the supervision of a researcher. The data collection protocol was approved by Fordham's Institutional Review Board. Users were asked to place one of our Android cell phones, running our data collection application, in their right front pants pocket and then to perform a set of activities for pre-defined periods of time, generally totaling 10 minutes each. Some users did not perform all activities due to physical limitations, and some activities (such as sitting and standing) were limited to only a few minutes because we expected that the data would remain fairly constant over time, which it, in fact, did. As users performed the activities, our application re-

References

1. [WISDM] Wireless Sensor Data Mining Project. Fordham University, Department of Computer and Information Science [Internet]. Bronx (NY). Available from: http://storm.cis.fordham.edu/~gweiss/ wisdm/.

2. Mantyjarvi J, Lindholdm M, Vildjounaite E., Makela SM, Ailisto H. 2005. Identifying users of portable devices from gait pattern with accelerometers. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing. ICASSP Pp. 973-976.

3. Gafurov D, Helkala K, Sondrol T. 2006. Biometric gait authentication using accelerometer sensor, Journal of Computers 1(7):51-59.

4. Gafurov D, Snekkenes E, Buvarp TE. 2006 Oct 29-Nov 3. Robustness of biometric gait authentication against impersonation attack," International Workshop on Information Security, OnTheMove Federated Conferences. France.

5. Mathie M, Celler B, Lovell N, Coster A. 2004. Classification of basic daily movements using a triaxial accelerometer.Medical & Biological Engineering and Computing: 42.

corded the accelerometer values every 50 ms (any faster and the data begins to repeat due to hardware limitations). When they had completed the set (walking, jogging, sitting, standing, ascending and descending stairs, and lying down), researchers copied the data from the application into our computers for future examination. Typical classification algorithms cannot interpret raw time series data;[14] rather, these algorithms classify examples. Thus we represent a period of data as a single example by transforming it into 43 descriptive features, (e.g. average values, time between peak values in the sinusoidal waves associated with repetitive steps, and descriptions of the distribution of values).

B. Modeling, Testing and Results

Our activity recognition task identifies seven activities from the accelerometer data: walking, jogging, climbing up and down stairs, sitting, standing, and lying down. These activities were chosen because they represent most of the activities smartphone users perform in the course of a day. The first step in evaluation is building classification models by feeding a standard classification algorithm training examples. These models are then tested for accuracy with new data. We find that generalized, impersonal models—those built from one set of subjects and tested on another—are, on average, 71% accurate. The advantage of this method is that a universal model can be downloaded and used by all. However, when a personal model is built from a single user's accelerometer data, the accuracy of the model on that user rises to an average 97%. This second scenario is akin to having application users train and personalize their devices before use. These results suggest that there are substantial differences in the way different people perform the same activities.

Our user authentication task uses the same data and techniques to identify the correct user from a pool of 36 initial users for whom we have data. Our results show that using only one sample containing 10 seconds of data, we can predict a user with about 72% accuracy. However, significantly better results can be achieved with more than 10 seconds of data. In order to identify a user, we use all of that user's data (typically 5-10 minutes worth) and make predictions on each sample within it, then choose the user who is most frequently predicted. This yields 100% accuracy for all 36 of our initial subjects. Thus, we are able to perfectly identify each of our 36 users based on their movements.

Conclusions and Future Work

The widespread use of sensor-packed mobile devices, including smartphones, tablet PC's, and gaming devices provides us with an unprecedented opportunity to study and develop applications for people's daily lives. User identification offers a broad range of possible applications. It can be used to provide device security and theft prevention. Identification can also be used to automatically personalize mobile device settings after identifying the current user of the device and his/her current activity. Applications that recognize activities and adapt phones as a result (such as selecting a certain playlist or sending calls to voicemail while running) can encourage healthy behavior. Moreover, the records of a user's activity can be tracked and reported over time, enabling health and fitness applications for users, and allowing people to see how sedentary they or their kids really are.

6. Brezmes, T., Gorricho, J.L., and Cotrina, J. 2009. Activity Recognition from accelerometer data on mobile phones. IWANN '09: Proceedings of the 10th International Work-Conference on Artificial Neural Networks. Pp. 796-799.

7. Cho, Y., Nam, Y., Choi, Y-J., and Cho, W-D. 2008. Smart- Buckle: human activity recognition using a 3-axis accelerometer and a wearable camera. In *HealthNet*.

8. Choudhury T, Consolvo S, Harrison B, LaMarca A, LeGrand L, Rahimi A, Rea A, Borriello G, Hemingway B, Klasnja P, Koscher K, Landay J, Lester J, WyattD, Haehnel D. 2008. The mobile sensing platform: An embedded activity recognition system. In IEEE Pervasive Computing; 7(2):32-41.

9. Gyorbiro N, Fabian A, Homanyi G. 2008. An activity recognition system for mobile phones. Mobile Networks and Applications; 14(1):82-91.

10. Krishnan N, Colbry D, Juillard C, Panchanathan S. 2008. Real time human activity recognition using tri-Axial accelerometers. Sensors, Signals and Information Processing Workshop.

11. Lester J, Choudhury T, Borriello G. 2006. A practical approach to recognizing physical activities. Lecture Notes in Computer Science: Pervasive Computing, 1–16.

12. Weiss GM, Hirsh H. 1998. Learning to predict rare events in event sequences. Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining. Menolo Park: AAAI Press. Pp. 359-363.

13. Witten IE, Frank E. 2005. Data mining: practical machine learning tools and techniques, 2nd ed. Morgan Kaufmann.

14. Miluzzo E, Lane N, Fodor K, Peterson R., Lu H, Musolesi M., Eisenman S, Zheng X Campbell A. 2008. Sensing meets mobile social networks: The design, implementation and evaluation of the CenceMe application. The 6th ACM Conference on Embedded Networked Sensor Systems: 337-350.

15. Jennifer R. Kwapisz, Gary M. Weiss and Samuel A. Moore. Activity Recognition using Cell Phone Accelerometers, Proceedings of the Fourth International Workshop on Knowledge Discovery from Sensor Data (at KDD-10), Washington DC, 10-18.