

My Smartphone Knows I am Hungry

Fanglin Chen, Rui Wang, Xia Zhou , Andrew T. Campbell
Dartmouth College
{chentc,ruiwang,xia,campbell}@cs.dartmouth.edu

ABSTRACT

Can a smartphone learn our eating habits without the user being in the loop? Clearly, the phone could use checkins based on location to infer that if you are in a cafe, for example, there is a good possibility you might eat or drink something. In this paper, we use inferred behavioral data and location history to predict if you are going to eat or not in the near future. These predictors could serve as a basis for future eating trackers that work unobtrusively in the background of your phone rather than relying on burdensome user input. We report on a simple model that predicts the food purchases of a group of undergraduate college students (N=25) using inferred behavioral and location data from smartphones. The 10-week study uses the dining related purchase records from student college cards as ground-truth to validate our prediction model. Initial results show that we can predict food and drink purchases with an accuracy of 74% using three weeks of training data.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous

Keywords

Food; human behavior dynamics; smartphone sensing.

1. INTRODUCTION

College students typically do not have healthy eating habits and as a result are at higher-risk of weight gain [5]. The basis of our work is to develop a predictive model capable of inferring the food buying habits of a student population. If we can build such a predictive model based on smartphone sensing data then we can provide just-in-time feedback to students about healthy food and drink choices.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

WPA '14, June 16 2014, Bretton Woods, NH, USA
Copyright 2014 ACM 978-1-4503-2825-8/14/06 ...\$15.00.
<http://dx.doi.org/10.1145/2611264.2611270>

Location (e.g., at cafe) is typically an excellent predictor of eating or drinking. However, many locations on a campus are multi-purpose and include restaurants, classrooms and arts facilities such as the Hopkins Center at Dartmouth College. We show that simply by using location you cannot easily infer eating from other activities without false positives or negatives. In addition, simply using the near instantaneous location label (e.g., Ramunto's Pizza) is too little too late in terms of potentially guiding a user to a better choice (e.g., Gym, followed by the Salad Hut). If we have predictive power to gain knowledge of what people might do in the near future based on the behavioral data and past location history we might be able to intervene before the user hits the bad food choice (e.g., fast food) in the first place.

In this paper, we propose a simple predictive model based on automatically inferred behavioral and location data from smartphones. The idea is that the phone is smart enough to determine if a user might eat or not in the near future without any burdensome input from the user. Such a predictive model could serve as a basis for novel interventions in the future or as a basis for implementing eating trackers.

We conduct a study of 25 undergraduate students at Dartmouth College over a single 10 week academic term. All the students live, eat and drink on campus. The students in the study ran the StudentLife app on their Android phones which inferred everyday activity (e.g., walking, stationary), sociability based on conversational data, and sleep, location, co-location, etc. To capture the participants food buying behavior, we collected the purchase history from their Dartmouth food cards at the end of the term. We use the sensing data on the phone to build a model that predicts the food buying habits captured by the Dartmouth food card. Our Initial results show that we can predict food and drink purchases with an accuracy of 74% using three weeks of training data. Note, our study only considers undergraduate purchasing behavior in a campus environment. We make no claims that our results present generalized behavior for a general population. We consider this future work.

2. RELATED WORK

There is little work on the prediction of eating habits using smartphone sensing data. In the participatory sensing community, researchers track users' eating behavior auto-

matically using cameras [11] or on-body sensors [2] with the goal of developing reliable dietary logging. Our work differs in that we focus on user’s food purchasing behavior before the user actually eats. This enables early interventions in the future to ensure user’s healthy food choice.

In the public health community, Hebden *et al.* [6] have developed applications that provide feedback based on population health guidelines including physical activity, intake of fruit and vegetables, and energy-dense takeouts. Unlike these applications that rely on the user to record their food intake, our system focuses on automating the process of tracking eating.

3. DATA COLLECTION

We collected a 10-week behavioral dataset for 25 participants from the StudentLife [12] study. The Institutional Review Board for the Dartmouth College approves this study. All participants are undergraduate students who live on Dartmouth College campus. Each student was given an unlocked Nexus 4 phone that runs the StudentLife app throughout the whole study. The app collects two types of data for our use: (i) inferred behavioral data from smartphones; and (ii) personal purchase history. The resulting dataset provides the basis for us to build and evaluate our prediction model.

3.1 Behavioral Data

The StudentLife app automatically infers user activity (*e.g.*, stationary, walking, running, driving, cycling), sleep duration, and sociability (*i.e.*, the number of independent conservations and their durations). It continuously runs as a background service on Android. The app also collects accelerometer, proximity, audio, light sensor readings, location, colocation (*i.e.*, the number of co-located Bluetooth devices), and application usage. The inferences and other sensor data are temporarily stored on the phone and then uploaded to our server when the user recharges their phone under WiFi.

The activity classifier [8] in the StudentLife app extracts features from the preprocessed accelerometer stream, then applies a decision tree to infer activity using a set of features. For each 10-min period, we calculate the ratio of non-stationary inferences. If the ratio is greater than a threshold, we consider the period active, meaning that the user is moving. We add up all 10-min active periods as the daily activity duration.

We consider the frequency and duration of conversations around a user as a measure of sociability. We infer that the user is “around conversation” using the privacy-sensitive audio and conversation classifiers [10]. The output of the classification pipeline captures the number of independent conversations and their durations.

Dartmouth College deploys WiFi APs across campus. We use the Wi-Fi scan logs (SSIDs, BSSIDs and RSSIs) on smartphones to locate the user. We obtain the detailed AP deployment information from Dartmouth network services. This allows us to associate a Wi-Fi scan log with a specific building. We then merge consecutive locations and partition a day into chunks, where the user stays in

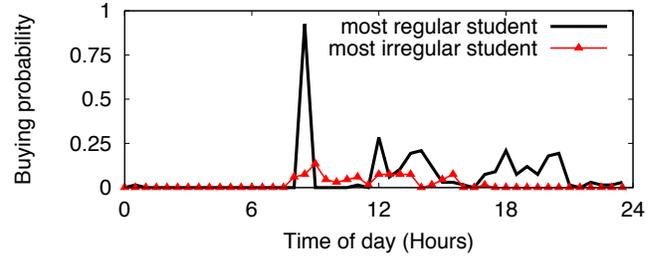


Figure 1: Time of day buying probability for the most regular student and most irregular student across 70 days.

the same building during each chunk. We assign a unique ID to each building for the classifier to process.

3.2 Purchase History

Each participant submitted their purchasing history captured by their Dartmouth card from early January to late May 2013. We only use the part of the history related to the StudentLife study. The Dartmouth card [1] is accepted at all college businesses for dining (soda, snacks and meals), entertainment (canoe, cabin rentals, ski, concert tickets, etc.) and services (printing, parking, laundry, etc.). We remove 8.65% of the original purchase history not related to our study, such as, for laundry, printing, mail, etc.

We use the rest of purchase history for the training and testing sets of the model. Each purchase record contains a corresponding location, cost and time stamp. Because all the undergraduate students are required to have a dining plan if they are enrolled in classes or living in campus housing, they use their Dartmouth card for most of their living expenses. Participants spent \$6.89 per purchase on average and made 2.81 purchases per day on average. We found 32 unique purchasing locations across Dartmouth campus in the purchase history for all the students.

4. PREDICTING FOOD PURCHASES

In what follows, we discuss a number of design challenges in building models for predicting food purchases for an undergraduate student body in a campus environment. We then present our model.

4.1 Design Challenges

Intuitively, location (*e.g.*, cafe) and time (*e.g.*, breakfast, lunch and dinner times) are excellent indicators of eating or drinking events. However, simply using these features is insufficient to predict food purchases for two reasons. First, many buildings on a campus are multipurpose and include restaurants, classrooms, and other facilities (*e.g.*, arts, entertainment, sports). Students can simply pass through or dwell a while in these buildings without making any purchases, or stay for longer periods of time to study or socialize for example. We study how likely a student might buy food when they are at one of these multipurpose building during meal time. Dartmouth dining defines meal times as follows: breakfast time during weekdays is

Features	Description
bld(c)	current building ID
bld(p)	last building ID
time(c) _{arr}	time of day upon arrival at the current building
time(p) _{arr}	time of day upon arrival at the last building
time(p) _{dept}	time of day upon departure at the last building
dur(p) _{conv}	total duration of all the conversations at the last building
freq(p) _{conv}	number of distinct conversation instances at the last building
ratio(p) _{act}	the percentage of time of non-stationary states in the last building

Table 1: List of candidate features.

from 7:30-10:30 am, lunch 11 am-3:00 pm and dinner 5-8:30 pm. We find that buying happens in only 26.7% of all the cases where students are at one of these multipurpose buildings during meal time. Clearly location and time alone can lead to significant prediction errors.

Second, most college students have irregular eating schedules. They often do not purchase food during meal time, and do not visit a routine cafeteria for meals. As shown in our dataset, only 24.9% of the participants purchased food during meal times. We plot the food purchase probability of the most regular student and the most irregular student in terms of their eating habits averaged over 10 weeks, as shown in in Figure 1. We see that the irregular student’s food purchase behaviors are almost evenly distributed during an average day. The regular student is more predictable during breakfast, but not for lunch and dinner. Only in 10% and 30% of the cases does the student buy lunch or dinner during common meal time periods. Therefore, time-based prediction will not work neither. We need a prediction model that addresses the time and location variability of food purchase.

4.2 Approach

To address these challenges, we identify characteristics from inferred behavioral data that are linked with user’s dining purchases. We extract features from these behavioral data and build a prediction model. Our goal is to predict if a student is going to purchase food at buildings where they can use their Dartmouth card. We refer to these buildings as "candidate buildings".

Specifically, our method consists of three steps: collecting training data, training the prediction model, and online prediction. In the training data collection stage, we collect features associated with the buying behavior, and purchase history as ground truth. In the training stage, we train a classification model that identifies predictive relations between the collected features and the purchase history. Finally, we use the prediction model obtained from the training phase to predict if the user is going to buy food after arriving at a candidate building. We formalize the prediction as a simple binary classification problem; that is, buying and not buying. We use classification and regression trees (CART) [3] as our classifier.

Table 1 summarizes a provisional set of features to predict food purchasing behaviors. We can divide these features into two categories: previous state features and current state features. Previous state features are calculated

using the behavioral data collected at the last building visited by a student. We propose these features because we hypothesize previous behavior states of students have causal effects on their choice of buying food. For example, after an workout at gym, students are more likely to buy a drink. Researchers in public health found that both physical activity and social behavior have influences on food choices [4, 7]. To capture physical activity (*i.e.*, how much students were physically active), we calculate the ratio of non-stationary labels during the time when a student visited the last building. To infer sociability, we use $dur(p)_{conv}$ and $freq(p)_{conv}$ to represent the conversation durations and the number of conversation instances at the last building, respectively. The current state features contain only the current building ID and the arrival time. Our model uses these features as input to compute online prediction results.

4.3 Prediction Model

In what follows, we discuss our prediction model, and the need for personalization and adaptation of the model. We personalize the model because of the diversity of purchasing behavior across students. We adapt the model because student’s purchasing behavior changes over the 10 week term.

4.3.1 Prediction Model

We use classification and regression trees (CART) [3] as the classifier. CART is able to handle both numerical and categorical data. In our case, the building name is categorical data that cannot be used directly in other classifiers. As a non-parametric decision tree, CART is commonly used to create a model that predicts the class based on the value of several features. We create the prediction model in the training phase, and use the model to predict food purchase.

The model is a set of prediction rules in the form of a tree. Each non-leaf node in the tree is a decision node which decides which node should go to next based on the value of a single feature. The algorithm traverse the tree from root to leaf based on feature values and each passing node’s rule until it reaches a leaf node. The leaf node gives the prediction label for the given feature input.

To build such a decision tree, we start with a single root node which contains all the training data instances. We recursively either divide the data at current node into subregions (*i.e.*, child nodes), or declare the node a leaf (*i.e.*,

assign a label). This process partitions the training data into sub data regions. Instances in each region shares the same label as well as similar feature values.

To split a node, the algorithm minimizes the impurity of data at child nodes. CART uses Gini [9] as the impurity measure. We choose Gini [9] because it measures feature importance in a decision tree. The Gini measure is defined as $i(t) = 1 - \sum_i p_i^2$, where p_i is the ratio of instances which have label i among all instances at the current node. The splitting stops when the impurity improvement of splitting is less than a threshold. At each node, the algorithm split the data in two based on a single feature that decreases the impurity. A feature’s Gini importance value is the sum of impurity decreases for all nodes that split based on this feature.

4.3.2 Personalization

We train a prediction model for each individual because different people have different buying behavior and location patterns. For example, different students may take different classes. The classes they take have different classrooms and schedules. Also, different students may choose to buy food at different campus stores. Thus, the location feature is not generic for all the students. A personalized model can accurately capture the characteristic of each student’s buying behavior. A generic model, however, is limited by the diversity of different students’ behavior.

4.3.3 Adaptation

A student’s buying behavior may change overtime. For example, as the term progresses, class workload increases. For example, some students may choose to stay longer at the library to study. To capture behavior change, we design a “dynamic adaptation mechanism”, where we use the most recent sensor data and purchase history to train a model that can reflect the most recent buying behavior, and recursively updated the prediction model over a fixed time interval.

We define a training period as a period of time in which the data collected will be used to train the prediction model. As we discussed in Section 4.2, there are two stages to produce the predictor: training data collection and prediction model training. The training data collection stage lasts for the length of a training period. At the end of the training period, a predictive model is trained using the data collected. Then, we use the model to predict the student’s buying behavior. At the same time, their sensing data and purchase history is collected. This prediction model expires after a period of time (e.g., 1 or 2 weeks) that is the same in length as a training period. A new prediction model is trained using the fresh data.

5. RESULTS

In what follows, we discuss the importance of different features, then present the performance of the classification model.

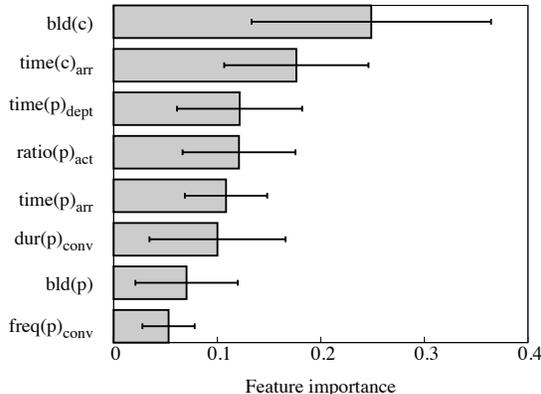


Figure 2: Ranking of feature importance.

5.1 Feature Importance

As discussed in Section 4.2, we use 8 features to predict if a student is going to buy food when they arrive at a building. Figure 2 shows the Gini importance for each feature. The top six important features are current building ID, arrival time at the current building, the departure time at the previous building, the activity ratio of the last building, arrival time at the previous building and the conversation duration at the previous building. Current states are more important than previous states. We also observe there are individual differences in term of feature importance. The variance of feature importance indicates if different users show similar behavioral patterns toward buying behavior. Within the top six important features, conversation duration at the previous building shows the most variance indicating that sociability is an important predictor for some individuals but not for others. Arrival time at the current location has the least variance, which indicates that it is a more generic predictor for the buying behavior.

5.2 Prediction Performance

5.2.1 Evaluation Method and Metrics

We identify 12260 cases where students arrive at one of the multipurpose buildings. In each case, we extract features describe in Section 4.2 as well as the ground truth which is a boolean value indicating if the student bought food during the visit to the building. Among all the cases, 3050 of them are related to buying. We evaluate and compare the prediction performance under three different model settings: generic model, personalized without adaptation and personalized with adaptation, as described in Section 5.2.2.

We use accuracy, precision and recall to evaluate the performance of the prediction model. Accuracy measures how well a binary classification test correctly identifies labels. It equals to $(tp+tn)/(tp+tn+fp+fn)$ where tp is the number of true positives, tn is the number of true negatives, fp is the number of false positives and fn is the number of false negatives. Precision measures the probability that a test case given positive label is truly positive. It is derived by $tp/(tp+fp)$. Recall measures the probability that a positive case can be identified by the classifier. It is derived by $tp/(tp+fn)$.

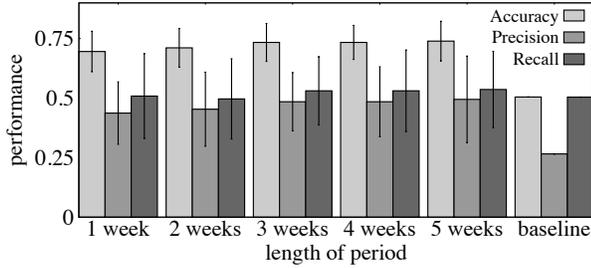


Figure 3: Prediction performance for the static training set.

5.2.2 Prediction Result

Using the above metrics, we now evaluate the performance of the different prediction model designs, including random labeling, a generic model without personalization, and a personalized model with and without adaptation.

Prediction Baseline. We use random guessing as the prediction baseline method, which guess if a user would buy or not buy food for a giving case. We run the baseline method on all 12260 cases in the dataset to get the prediction performance measured by accuracy, precision and recall. The results shows the accuracy of the baseline method is 50.5%, precision is 26.6% and recall is 50.4%.

Generic Model. In this setting, the dataset is randomly partitioned into a training set and test set. We use 10-fold cross validation to evaluate the prediction performance. The accuracy is 68.6%, precision is 42.1% and recall is 49.3%. The generic model outperforms the baseline in terms of prediction accuracy and precision. But it is 1.1% worse than the baseline in terms of recall. This indicates that the model can predict the label of each case more accurately than random guessing, and the predicted positive cases are much more likely to be correct. However, it cannot identify positive cases from negative cases as efficiently as random guessing.

Personalized Model without Adaptation. In this setting, we partition the dataset by student. We segment a student’s data into training set and test set according to time. We determine how much data to use as the training set and how much for the test data, then segment the data by time. For example we use the first 2 weeks data as our training set and the rest as our test set. This setting models a real life scenario in which the prediction model learns user’s buying behavior first, then we use the learned model to predict future buying behaviors.

We calculate the mean and standard deviation of the evaluation metrics for each student’s test set. Figure 3 shows the performance for different sizes of the training set in terms of time. We observe the overall performance is significantly better than the baseline. As the size of the training set increases, the performance in term of accuracy, precision and recall also increase. We achieve the best performance when the first 5 weeks of data is used for training. The accuracy for this setting is 73.9%, precision is 49.5% and recall is 53.6%.

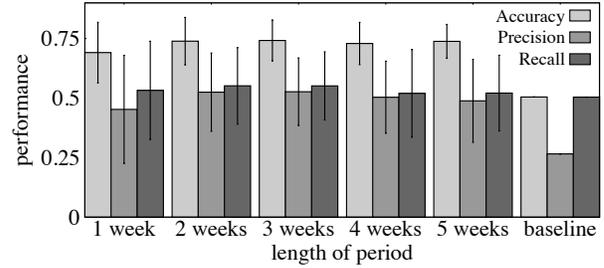


Figure 4: Prediction performance for different adaption periods.

We can see from the results shown in that the model can identify positive cases from negative cases more efficiently compared to the generic model and random guessing. Prediction accuracy and precision is further improved.

Personalized Model with Adaptation. We test the model with five adaptation periods. The results inform us about the optimal adaptation period for the dataset. Each student’s data is partitioned into segments. Each segment’s length in term of time is the length of a training period. We recursively use one data segment as the training set to predict the consecutive segment’s buying behavior. The results are shown in Figure 4. It shows the classifier significantly outperforms the baseline in term of accuracy and precision. The recall rate is 3.1% higher than the baseline. It indicates that when the model predicts a student is going to buy food, it is more likely that it is true compared to random guessing.

In terms of the performance of the different adaption periods, the prediction performance improves as the period increases to 3 weeks, the accuracy increases from 69.1% to 74.2%, precision from 45.2% to 52.7%, recall from 53.2% to 55.1%, then drops for 4 and 5 week periods, accuracy drops to 73.8%, precision to 48.9% and recall to 52.1%. The change of performance is more significant in term of precision and recall.

Summary. We can see from the results shown in Figure 4 that personalization significantly improves the prediction performance as one would expect; the accuracy, precision and recall improves by 5.3%, 7.4%, 4.3% respectively. The precision and recall further improve by 4.2% in precision and 1.5% in recall by using adaptation. This results confirm our hypothesis discussed in Section 4.3 that different students have different buying behavior patterns that change overtime. Our personalized and dynamically adapting model can capture individual differences and changing buy behavior over time.

6. CONCLUSION

In this paper, we started by asking if smartphones can predict our food purchase behaviors. We reported on a simple predictive model of food purchases. Specifically, we design this model to predict whether a smartphone user would choose to buy food in a near future or not. We studied a group of undergraduate students (N=25) who collected inferred behavioral and location data from their smartphones over 10 weeks. Our initial results show that

we can predict food purchases with an accuracy of 74% using three weeks of training data. Because of the unique characteristic this college student dataset, this model can form the basis of new interventions and trackers that can help reduce unhealthy eating behavior among college students. In order to generalize our work, we intend to explore more features for prediction of more types of food purchases including purchase costs, purchase type, total number of daily purchase instances. Also, we will look at the food purchase behavior of the office workers because they shared similar life pattern with students.

7. REFERENCES

- [1] Dartmouth dining service. <http://www.dartmouth.edu/dining/plans/>.
- [2] Amft, O., and Tröster, G. Recognition of dietary activity events using on-body sensors. *Artificial Intelligence in Medicine* 42, 2 (2008), 121–136.
- [3] Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. *Classification and regression trees*. CRC press, 1984.
- [4] Feunekes, G. I., de Graaf, C., Meyboom, S., and van Staveren, W. A. Food choice and fat intake of adolescents and adults: associations of intakes within social networks. *Preventive medicine* 27, 5 (1998), 645–656.
- [5] Flegal, K. M., Carroll, M. D., Ogden, C. L., and Johnson, C. L. Prevalence and trends in obesity among us adults, 1999-2000. *Jama* 288, 14 (2002), 1723–1727.
- [6] Hebden, L., Cook, A., van der Ploeg, H. P., and Allman-Farinelli, M. Development of smartphone applications for nutrition and physical activity behavior change. *JMIR Research Protocols* 1, 2 (2012).
- [7] Lowry, R., Galuska, D. A., Fulton, J. E., Wechsler, H., Kann, L., and Collins, J. L. Physical activity, food choice, and weight management goals and practices among us college students. *American Journal of Preventive Medicine* 18, 1 (2000), 18–27.
- [8] Lu, H., Yang, J., Liu, Z., Lane, N. D., Choudhury, T., and Campbell, A. T. The jigsaw continuous sensing engine for mobile phone applications. In *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems* (2010), ACM, pp. 71–84.
- [9] Menze, B. H., Kelm, B. M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., and Hamprecht, F. A. A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC bioinformatics* 10, 1 (2009), 213.
- [10] Rabbi, M., Ali, S., Choudhury, T., and Berke, E. Passive and in-situ assessment of mental and physical well-being using mobile sensors. In *Proceedings of the 13th international conference on Ubiquitous computing* (2011), ACM, pp. 385–394.
- [11] Reddy, S., Parker, A., Hyman, J., Burke, J., Estrin, D., and Hansen, M. Image browsing, processing, and clustering for participatory sensing: lessons from a dietsense prototype. In *Proceedings of the 4th workshop on Embedded networked sensors* (2007), ACM, pp. 13–17.
- [12] Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D., and Campbell, A. T. StudentLife: Assessing mental well-being, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM Conference on Ubiquitous Computing* (2014), ACM.