Sequencing the Dietary Exposome with
Semi-Automated Food Journaling Techniques

Edison Thomaz
School of Interactive Computing
Georgia Institute of Technology
ethomaz@gatech.edu

Abstract: Despite our understanding of the impact of lifestyle on human health, we lack tools and techniques that capture individuals' behavioral exposures such as diet, sleep and exercise over time. My current work focuses specifically on capturing eating habits, where I am currently exploring semi-automated food journaling approaches.

Keyword: Diet

Prepared for the HCIC 2013 Workshop, June 23rd to June 27th 2013

Introduction

Over the last 10 years, there has been much excitement about the potential of our newly-acquired knowledge of the human genome towards understanding human health and the underlying cause of disease. However, as we now know, genetics appears to account for only 10% of diseases, whereas the remaining 90% seem to be attributed to environmental factors, the so-called "exposome". Unfortunately, our understanding of the impact of environmental and lifestyle factors on human health is very limited. The main reason for this is because we lack tools and techniques to collect detailed longitudinal data characterizing one's lifestyle (e.g. sleep, exercise) at scale.

One of the key factors affecting an individual's health is diet. In 2008, one third of all adults in the U.S were overweight or obese, with other countries observing similar trends [13]. It is believed that an effective method to monitor eating habits could provide insights into this seriously growing problem. One of the fundamental problems in characterizing eating habits is that there is not an efficient way to collect dietary information that is objective, ecologically valid and that does not pose a major burden on individuals. Today, the “state-of-the-art” in personal food logging lives within the domain of mobile phones and mobile phone apps. There are a myriad of applications that let users take photos and notes of their meals, some of which go a step further and even display the nutritional value or the health score of a meal through crowdsourcing techniques, such as MealSnap. The key challenge with these applications is that people need to remember to use them, which proves to be particularly hard to do in the long term. Additionally, even when people remember to use these applications, there is a cost associated with fetching a smartphone, unlocking it, launching an app and taking a photo or typing notes. It is inevitable that even the most engaged users might forget to log a snack or meal occasionally, or grow weary of dutiful logging over the long run. The truth is, these applications are simply not practical enough for sustained use.

Research in the area of food recognition dates back to the 1980s when researchers tried to detect chews and swallows using oral sensors in order to measure the palatability and satiating value of foods [25]. Other sensor-based techniques involve detecting eating and drinking actions from inertial sensors attached to the upper and lower arms [1] and monitoring caloric intake using on-body or mobile phone-based sensors {Chen:wl}. While sensor-based approaches are able to derive information directly from body motions, they are obtrusive and do not make use of the valuable visual cues.

Recently, the ubiquity and popularization of a number of technologies from sensors to wearable devices has made it possible to envision systems that completely automate the capture of dietary intake. In practice, however, these systems have also been extremely difficult and complex to implement. Analyzing food images with computer vision algorithms, and addressing privacy concerns are some examples of tasks that have been explored with promising results by researchers but that need to be further developed to be deployable in real-world settings. Furthermore, even if the capture of dietary intake could be fully automated, it might not desirable in many cases since it would exclude individuals from the process of data collection. This is because it has been shown that self-monitoring contributes to positive health outcomes not only in terms of weight-loss but well being in general. In other words, getting individuals to be engaged in the logging of their dietary intake is important.

In my current line of work I am exploring the space in-between manual and fully automated food journaling with the goal of leveraging the best characteristics of both approaches. I refer to this technique as semi-automated food journaling. I rely on sensing technologies, wearable devices, and interactive machine learning techniques to infer patterns and instances of eating activity and subsequently prompt individuals for details about these eating activities.

There are three research threads within this work:

1. Aggregation of human activity-centric sensor data

2. Inference of eating activity patterns

3. Design of a wrist-based device to improve ability to complete a food journal

Aggregation of Human Activity-Centric Sensor Data

Thanks to advances in sensing and mobile technologies over the last decade, researchers have employed a wide variety of sensors to automatically infer many aspects of human activity [14] [17] [15] [9]. Recently, wearable devices that leverage sensor and communication technologies to log physical activity have advanced beyond research labs to become very popular in the consumer market. Some of these devices include the Fitbit, the Nike FuelBand and the Garmin Forerunner.

Despite these positive developments, many important dimensions of an individual's everyday lifestyle remain outside the reach of current sensing technologies. This is due in large part to the complexity of certain types of activities, such as eating. Some of the characteristics of an eating activity that would be desirable to capture include (1) when eating is taking place, (2) what is being consumed, and (3) how much is being consumed. It is not possible to capture the totality of an eating activity with one sensor automatically. However, by aggregating data from multiple sensing sources and incorporating additional lightweight sensing modalities, we believe it is possible to recognize an individual's eating activity in the moment based on a priori sensor values, and also build models that reflect an individual's eating patterns over time.

For example, by examining an individual's location from GPS data (e.g. close to the office), her amount of physical activity (e.g. little movement), day of week and time of day (e.g. Tuesday at 1PM), and on-body acoustic sensing in mouth, neck and throat (e.g. indicating chewing, drinking and speaking) [1, 16, 21, 28] [25], it is highly likely that the individual is having lunch. Confidence in this inference could be raised even more if a lunch event could be observed in the individual's calendar for 12:30PM.

Once a meal activity has been identified, several courses of actions might be pursued. An automatic trigger could be sent to a wearable camera to take a picture of the food [18] [26], the individual could be nudged to add an entry to a food logging mobile application [6] [2], or a text message could sent to the individual later in the day requesting more details about the meal. As it becomes evident, the identification of when a meal takes place is the centerpiece of a number of strategies for automatic and semi-automatic food journaling.

The first step towards this vision involves building an aggregator for multiple sensor streams. I am building an aggregator that accept single-point and multi-point data from two types of sources, devices and services. Single-point and multi-point refer to the number of data points that a source writes to the aggregator at any one time (i.e. one data point at a time vs. multiple data points at a time). The aggregator's database schema will be created to be as simple as possible to use and understand, but still able to store data from a variety of sources, as discussed below.

Inference of Eating Activity Patterns

Traditionally, activity recognition systems are implemented using supervised machine learning (ML) techniques [4, 27]. Using these kinds of algorithms, which include Neural Networks, Support-Vector Machines (SVM) and Decision-Trees, building an activity classifier requires a training set with annotated data. In most cases, however, compiling such a training data set proves to be a challenge. This is because annotating data while performing everyday activities is a time-consuming and error-prone task. Moreover, given individual differences and population variability, one model trained to recognize tasks performed by one person may not recognize tasks performed by others. In other words, models built this way, and in particular the data used to construct these models, do not generalize well.

Alternatively, researchers interested in discovering the structure of people's routines, so called "life patterns", have relied a number of unsupervised ML techniques. Some of these approaches include methods for finding discontinuous and varied-order activity patterns and computing the principal components of an individual's behavioral data [5, 8, 11, 22]. One of the challenges of these unsupervised approaches is the amount of data required. Even though they are not supervised ML techniques, and thus do not require a labeled training set, they still require a substantial amount of data. For example, Eagle and Pentland reported being able to obtain 95% accuracy in cluster separation using Expectation-Maximization after training their model with one month of data from several subjects [10]. Another consideration is that once patterns have been detected, it is critical to learn what activities the patterns refer to. Interactive machine learning techniques, where end-users provide labels or features to guide the process of learning, can be used towards this end.

A practical approach for inferring eating activities patterns from sensor data should address the following three questions:

1. How to predict specific life patterns (i.e. eating) from aggregated sensor data?

2. How do we obtain labels and information from end-users with regards to particular activity patterns in a way that does not pose a burden and is not perceived as disruptive?

3. How to infer life pattern (e.g. having lunch) in real-time from sensor data and previously built life patterns models?

Predicting life patterns (i.e. eating) from aggregated sensor data

Routine at all temporal scales characterize aspects of human life for many individuals [10]. The first step towards predicting specific life patterns involves clustering sensor data streams across time. In previous work, we have identified that the amount of physical activity observed by an on-body accelerometer and also an individual's location can be used as predictor for eating activity. I also expect that on-body acoustic sensing might be a useful feature in this scenario [1, 21, 28].

After several days or weeks, depending on the confidence desired to predict patterns, a set of clusters will be available for each day. Translating these cluster sets into one set that corresponds to an individual's habitual activities can be achieved by coalescing the clusters on a timeline, through time-alignment. We call these coalesced clusters life pattern clusters.

Obtaining life patterns labels

Once life pattern clusters (LPCs) have been identified, the next step involves obtaining labels for them. Querying individuals through an SMS messaging interface might be one way to achieve this. To avoid overburdening individuals with too many messages, only a few queries should be submitted per day.

Inferring eating activity from sensor data

LPCs correspond to patterns observed from low-level sensor data. Given labeled LPCs, it becomes possible to compile a training set that can be used for building an LPC classifier using supervised ML techniques. These classifiers can then be applied towards the inference of eating activity in real-time.

Wrist-based device to improve ability to complete a food journal

Nowadays, one of the popular ways to keep track of a food journal is through a mobile phone application. The use of mobile apps is compelling because most people already carry their mobile devices with them when outside the home. Additionally, there are many food log applications to choose from, suiting a variety of personal journaling styles. However, in spite of these factors, adherence to mobile food logging is often short-lived and tied to temporary health goals (e.g., weight loss). This is caused in large part by the effort required in remembering to log every eating activity and then taking the manual steps required to do so, which include taking the mobile phone out of pocket, unlocking it, finding the food journaling application, etc.

With the emergence of devices such as Google Glass and the Memoto camera, it becomes possible to devise systems that capture people's eating activities completely automatically through first-person point-of-view images. In practice, there are three key downsides to this approach, (1) lack of control and privacy concerns when images are taken automatically, (2) large number of images to analyze, and (3) societal norms and pressure against the use of such wearable devices in public.

Over the last 15 years, researchers have been exploring the space of wearable devices and micro-interactions to enable new kinds of experiences and facilitate the completion of tasks [3, 7, 12, 19, 20, 23, 24].

To facilitate the process of food logging in real world settings, I suggest a new wearable wrist-based camera device that I call "WristPhoto". WristPhoto leverages micro-interactions to remind individuals to document their meals (i.e. by taking a photo of their food), and making it effortless to do so. The device should satisfy three important conditions:

Remind individuals to take snapshots during eating activities: A smartphone-grade vibrator motor can be integrated into the device and activate for a very short period of time whenever the activity classifier running in WristPhoto recognizes that an eating activity is taking place.

Require minimal access time: Access time is of the key aspects that differentiate the WristPhoto from a mobile phone when it comes to taking a photo for food journaling purposes. Shooting a photo with the device might be as simple as pointing to an object, or a plate of food, and performing a quick and intuitive hand gesture. A sensor in the device could recognize the gesture and instruct the camera to take a snapshot. We expect that the access time of taking a photo to be within 1-2 seconds.

Designed in a socially-acceptable form factor: Lately a number of products such as the Nike FuelBand and the Jawbone Up have popularized the wristband form factor for activity tracking. The WristPhoto will also sit on the wrist, and in the envisioned final form, have the same aesthetics as these other devices.

References

[1] Amft, O. et al. 2005. Analysis of chewing sounds for dietary monitoring. *Proceedings of the 7th International Conference on Ubiquitous Computing (Ubicomp 2005).*

[2] Andrew, A.H. et al. 2013. Simplifying Mobile Phone Food Diaries: Design and Evaluation of a Food Index-Based Nutrition Diary. *Proceedings of the International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth 2013).*

[3] Ashbrook, D. 2010. Enabling mobile microinteractions. *PhD Thesis*. (Jan. 2010).

[4] Bao, L. and Intille, S. 2004. Activity recognition from user-annotated acceleration data. *Pervasive Computing*. (2004), 1–17.

[5] Begole, J.B. et al. 2003. Rhythm modeling, visualizations and applications. (Nov. 2003). *Proceedings of the 16th annual ACM symposium on User interface software and technology (UIST2003).*

[6] Bentley, F. and Tollmar, K. 2013. The Power of Mobile Notifications to Increase Wellbeing Logging Behavior. *Proceedings of the ACM SIGCHI International Conference on Human Factors in Computing Systems (CHI2013)*.

[7] Christian Loclair, S.G.P.B. 2010. PinchWatch: A Wearable Device for One-Handed Microinteractions. *Proceedings of MobileHCI 2010.*

[8] Clarkson, B.P. *Life Patterns*. PhD Thesis. (2002)

[9] Consolvo, S. et al. 2008. Activity sensing in the wild: a field trial of ubifit garden. *Proceeding of the ACM SIGCHI Conference on Human factors in Computing Systems (CHI2008).*

[10] Eagle, N. and Pentland, A. 2006. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*. 10, 4 (2006), 255–268.

[11] Eagle, N. and Pentland, A.S. 2009. Eigenbehaviors: identifying structure in routine. *Behavioral Ecology and Sociobiology*. 63, 7 (2009), 1057–1066.

[12] Harrison, C. et al. 2010. Skinput: appropriating the body as an input surface. *Proceeding of the ACM SIGCHI Conference on Human factors in Computing Systems (CHI2010).*

[13] Kimokoti, R.W.R. and Millen, B.E.B. 2011. Diet, the global obesity epidemic, and prevention. *YJADA*. 111, 8 (Aug. 2011), 1137–1140.

[14] Lane, N. et al. 2010. A survey of mobile phone sensing. *Communications Magazine, IEEE*. 48, 9 (2010), 140–150.

[15] Lin, M. et al. 2012. BeWell+: Multi-dimensional Wellbeing Monitoring with Community-guided User Feedback and Energy Optimization. *Proceeding of the Wireless Health Academic/Industry Conference (Wireless Health'12)*. (2012).

[16] Lopez-Meyer, P. et al. 2012. Automatic identification of the number of food items in a meal using clustering techniques based on the monitoring of swallowing and chewing. *Biomedical Signal Processing and Control*. 7, 5 (Sep. 2012), 474–480.

[17] Lu, H. et al. 2010. The Jigsaw continuous sensing engine for mobile phone applications. *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*. (2010), 71–84.

[18] Martin, C.K. et al. 2008. A novel method to remotely measure food intake of free-living individuals in real time: the remote food photography method. *British Journal of Nutrition*. 101, 03 (Jul. 2008), 446.

[19] Mistry, P. et al. 2009. WUW - wear Ur world: a wearable gestural interface. *CHI EA "09: CHI "09 Extended Abstracts on Human Factors in Computing Systems*. (Apr. 2009).

[20] Nanayakkara, S. et al. 2013. EyeRing: a finger-worn input device for seamless interactions with our surroundings. *Proceedings of the 4th Augmented Human International Conference (AH13)*.

[21] Passler, S. and Fischer, W. 2011. Acoustical method for objective food intake monitoring using a wearable sensor system. (2011), 266–269. *Proceedings of the International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth 2011).*

[22] Rashidi, P. and Cook, D. 2010. Mining and monitoring patterns of daily routines for assisted living in real world settings. *Proceedings of the 1st ACM International Health Informatics Symposium*. (2010), 336–345.

[23] Rekimoto, J. 2001. GestureWrist and GesturePad: unobtrusive wearable interaction devices. *Proceedings. Sixth International Symposium on Wearable Computers*. (Oct. 2001), 21–27.

[24] Starner, T. et al. 1998. A wearable computer based american sign language recognizer. (1998), 84–96. *Assistive Technology and Artificial Intelligence.*

[25] Stellar, E. and Shrager, E.E. 1985. Chews and swallows and the microstructure of eating. *The American journal of clinical nutrition*. 42, 5 (1985), 973–982.

[26] Sun, M. et al. 2010. A wearable electronic system for objective dietary assessment. *Journal of the American Dietetic Association*. 110, 1 (2010), 45.

[27] Van Kasteren, T. et al. 2008. Accurate activity recognition in a home setting. *Proceedings of the 10th international conference on Ubiquitous computing* (Ubicomp 2008).

[28] Yatani, K. and Truong, K.N. 2012. BodyScope: A Wearable Acoustic Sensor for Activity Recognition. (2012). *Proceedings of the International Conference on Ubiquitous Computing (Ubicomp 2012)*.