Reflections of Everyday Activities in Spending Data

ABSTRACT
In this paper we show that financial information can be used to sense many aspects of human activity. This simple technique gives people information about their daily lives, is easily accessible to many at no extra cost, requires little setup, and does not require the manufacture of any external devices. We will focus on how financial data can be used to show users where they spend their time, when they accomplish certain habits, and what the impact of their activities is on the environment. We validate our idea by implementing three demonstration applications intended for personal use. Finally, this paper discusses the privacy implications of sensing activity using financial data.

Author Keywords
Ubiquitous Computing, Context-Aware Computing

ACM Classification Keywords
H5.2 [Information interfaces and presentation]: User Interfaces- Graphical user interfaces.

INTRODUCTION
Researchers have been touting the potential benefits of having a complete record of daily activity since Vannevar Bush first proposed the Memex, a device that would record important information about an individual’s activities and support later retrieval of that information [2]. Recent instantiations of this vision, termed life-logging technologies, have demonstrated the value of passive approaches to capturing a record of a person’s activities [4,6]. This data can then be used to support a variety of activities such as cueing memories [6], or creating a travelogue [4]. Yet these applications currently require both a significant investment in infrastructure and access to unusual hardware.

In this article, we propose using a ubiquitously available, inexpensive source of data to expand the information available about an individual’s activities. In today’s world, almost everything we do has some sort of financial impact. Financial data may show a warped and incomplete record of our activities, but it is nonetheless more complete than many other digital records currently available.

Financial data is a plentiful and readily available data source that anyone can use to learn about his/her life. Personal financial data has recently become easily available in a consistent form through online banking and credit cards. Now, online applications to help people manage this information are blossoming (e.g. Mint.com, Wesabe.com).

Financial transactions typically have a user or system-provided tag, a date, an amount, and a merchant associated with them. Tagging of financial data is partly or fully automated by today’s tools and enables many powerful features for users. We leverage the meta-data encoded in tags and combine it with information about dollars spent, date, and merchant to extract additional context, such as the frequency of certain habits and the carbon emissions of a person's activities.

We validate our idea by describing three applications that show individuals reflections of their financial data (all available at http://finance.notjulie.com). To implement them, we used a publicly available API of one of these applications (Wesabe) as our financial data source. Each of our applications aims to illustrate how financial data can be used to tell users more about themselves than just what they spend their money on:

1. Snoopy is a simple end-user programming system that shows users when they did certain habits and visualizes how those habits change during the week and over time.
2. Co2Green is a carbon footprint calculator that gives people an idea of what the impact of their activities was on the environment. Co2Green converts the dollar value of certain tagged transactions into carbon emissions and displays carbon emissions over time.
3. Personal Bubble/Errand Planner combines merchant location information with financial information to create a visualization that shows people where they spend their time. It also suggests the most efficient order for completing daily errands (requiring the least travel distance).

It is important to note that there are several drawbacks to using financial data to detect human activity. First, financial data is currently not available in real-time, limiting the set of possible applications. Even though this is likely to change in the future, our applications demonstrate the value of our approach even without real-time data. Second, there are sources of inaccuracy that may require human interpretation. For example, not all purchases a person makes are for his/herself. Third, a user may have to change his/her choice of tags to make better use of our applications. This is unlikely to be a problem with newly tagged data, but might affect the usefulness of past data. Finally, using financial data to track activity is a privacy threat, though one that is already ongoing. Our applications at least give users some of the power already in the hands of their banks.
and credit card companies. Despite these drawbacks, there is a great deal of useful information that users can obtain about their lives by looking at financial data.

BACKGROUND

The analysis of personal financial data is a well-examined field, focused on understanding and predicting purchasing habits. Personal finance managers such as Quicken, Wesabe.com, and Mint.com use personal financial data to extract and display information such as the value of users’ purchases in various categories. While there are interesting uses of these tags when looked at a large scale [1], the most familiar visualization is the pie-chart which shows users their top spending categories.

Data mining can be used to create purchasing models for individuals using transactional data. For example, Cadez, et al., examined ways to use a mixture model in order to predict individuals’ future purchases given a large set of purchasing data [3]. Such models have been used to recommend products or give other suggestions. For example, Amazon.com recommends products similar to items customers have bought in the past and Wesabe.com recommends similar vendors users may want to visit.

Moving outside of a focus on purchasing, Intuit, Inc. (the maker of Quicken) has recently received a patent to use transaction information to suggest or represent social information in a social network [11].

In summary, most past uses of financial data have focused on making predictions about purchasing or visualizing information about purchasing. Our work demonstrates the potential to move outside of purchasing and leverage information about human activities that is encoded in financial data. The remainder of this paper describes three applications demonstrating the range of information that can be extracted from financial data.

SNOOPY: VISUALIZING HABITS

One of the advantages of financial data is the ease with which users can gather a year or more of historical information for analysis. But lists of thousands of transactions are difficult to interpret, and simple visualizations that show spending by category may not capture everything that a user wishes to know about what happened when.

Snoopy allows users to define habits in terms of their transactions, shows which days of the week they do certain habits, and visualizes those habits over time. Many habits such as buying coffee, gambling, or shopping too much can easily be derived by looking at tagged purchases.

Application Description

As with all of our applications, Snoopy was written in JavaScript and implemented on top of the Wesabe API which allows the secure download of a user’s transactions. After a user loads data from Wesabe, he/she sees a histogram of tagged transactions binned by weekday. The user can easily explore particular tags by selecting any tag or set of tags and viewing a histogram of just that subset of the data (Figure 1). The user can also view how his/her transactions change over time on a timeline graph. Both graphs are implemented using Google’s Visualization API.

Users can also define their own habits by completing a simple sentence of the form: “I [X] if I spent [More than | less than | exactly] [Y] on purchases with the tag [RegExp].” “Users can also define habit as occurring once for every [Y] spent on purchases with the tag [RegExp]. The habit X is then defined, and users can explore one or more habits using the same interface as they use to explore tags.

Validation

We tested Snoopy on two data sets belonging to two of the authors. One data set is about 14 months in size, the other about 6 months in size. Our validation is informal, but demonstrates that Snoopy can be used to explore personally relevant activities with real data: One author defined exercise habits in terms of use of a parking structure (rather than biking to work) and taking a yoga class. The other author tracked a planned reduction in coffee drinking (by tracking dollars spent with the tag ‘coffee’) and learned that she almost never drinks coffee on Mondays (Figure 1). She also used snoopy to compare how much she eats out as opposed to eating in (Figure 2). “Eating out” was defined as a purchase with the tags ‘fast-food | restaurant | lunch’.” “Eating in” was defined as happening 1 time for every $10 spent on transactions with the tag ‘groceries’. Figure 2 shows that the author used to eat out more than she ate in, but has begun eating in more.
Discussion
Snoopy illustrates that it is possible to define habits in terms of financial transactions, and shows that financial data can be used to show people when they do things, and how their habits change over time. While our goal was to demonstrate a proof of concept, our use of the system illustrated the need to support end-user programming: the best way to define a habit is best known by the person engaged in it. There are well known end user programming techniques to specify complex definitions for habits that Snoopy could eventually leverage (e.g. [10]).

CO2GREEN: IMPACT ON THE ENVIRONMENT
While financial data can be used to track activities with a user’s help, it is more appropriate for tracking the impact of activities. In particular, there is a strong connection between dollars spent and environmental impact [7]. While at least one company (www.footprintscanner.com) will compute a carbon footprint from financial data to large corporations, their approach is neither automated nor tailored to the needs of individual users.

The current state of the art in tracking carbon emissions is fairly limited: individuals can take a survey that aims to calculate their carbon footprint. Different calculators vary in their level of detail, and there is a lack of consistency across calculators [2]. Consistency is further compromised because users often do not know answers to questions such as how many gallons of gas their car consumes in a year. Financial data provides a way to more systematically track carbon emissions that requires little extra effort for the user. It also enables fine-grained detail (weekly or monthly views of impact) that can lead to new insights.

Application Description
Co2Green calculates the carbon emissions for every relevant transaction, and then displays these emissions over time (http://calculator.stepgreen.org). We used national average data from for the year 2008 to convert spending dollars for car fuel, natural gas and electricity into carbon emissions[6, 9, 13]. For air travel we ask users to specify if flights were short, medium, or long, and converted this to emissions using numbers typical of existing calculators.

Once a user loads his/her tagged data and specifies which tags indicate purchases of auto fuel, natural gas, electricity and airplane flights, he/she sees a bar graph of the calculated emissions split up by category (fuel, natural gas, airplane, electricity) over the last 12 months (Figure 3). Instead of getting a single number as they would from a standard calculator, users are able to see how their emissions change by week, month or year. This provides insight into how a user’s emissions change based on season, temperature, and circumstance.

Validation
We validated Co2Green in a small user study. Users were asked to use Co2Green and a typical carbon footprint calculator from Zerofootprint.net. Afterwards they filled out a brief survey comparing the two. To make the results more comparable, we modified the zerofootprint.net calculator to only ask questions that could be answered using financial data. Six users completed the study, their median age was 31-40 and their median income was 20-30k. All were employed full time. Users reportedly spent about 5 minutes modifying the tags in their data in preparation for using co2green, and all had used Wesabe before.

Most users did not feel that Co2Green was accurately representing their footprint. Perhaps as a result of this, they rated it as equal to or slightly lower than Zerofootprint on all usability measures. Users told us that their estimation of accuracy was based on (1) the specificity of the data being used (both calculators performed well on this metric) and (2) whether or not certain financial data was missing or confounded by a user's tagging scheme. Users liked the monthly breakdown of emissions provided by Co2Green.

Discussion
The carbon footprint calculator illustrates how financial data can be used to show users the impact that their activities have on the world. We plan to add support for describing emissions by answering questions similar to those used by zerofootprint, and by using rules similar to those in Snoopy. This will allow users to better account for complex tagging issues like those raised in our user study.

PERSONAL BUBBLE AND ERRAND PLANNER: VISUALIZING WHERE
From the extensive use of location-aware technologies such as GPS, it is clear that knowing where people are is very important. While location isn’t normally included in financial data, it can be added easily using similar predictive techniques to tagging (i.e. remembering past values for each merchant), or potentially by comparing merchant information with sensed location. Financial transactions provide additional context about spending not normally available to sensors.

The personal bubble visualizes where individuals spend most of their money (Figure 4). This could be used as a tool to motivate people to reduce their carbon emissions by travelling less, but it does not provide direct suggestions for how to do this.
The errand planner suggests how errands could be combined by finding the most efficient order (shortest distance) for accomplish multiple purchases in one day.

**Application Description**

After a user loads their financial data he/she sees which stores he/she visits grouped by day. The user can select which of these groups are errands and indicate what the start and end address is when he/she visits these stores (usually his/her home). The errand planner displays the route with the shortest travel distance on a map (Figure 5). Our implementation enumerates all possible orderings, checking distance using Google Maps API.

**Validation**

We tested our concept on seven participants’ data. We found that 64% of all days when purchases were made had multiple purchases. The largest number of purchases users made in one day was 8, and the average number run was 1.8. Our algorithm works quickly enough on such small enumerations to not be a problem. The most efficient route based off of the seven errands the first author identified was, on average, 16% more efficient than the least efficient.

**PRIVACY CONCERNS**

Privacy, an important concern in any application that senses human activity, is particularly pertinent when working with financial data. The applications described here are intended for personal use, and could be enhanced by revealing some general and anonymous information about a user’s activities without compromising privacy. It is our hope that the benefits of examining financial data outweigh the potential negatives.

Even so, the idea of sensing human activity using financial data is inherently frightening to many people. As our background section shows, much is already being extracted from financial data. By showing these techniques in a public venue, we are educating people—telling them what external agents may be doing with their data and letting people access this data themselves.

**CONCLUSIONS AND FUTURE WORK**

We have demonstrated the potential for using financial data to sense human activity by the showing when, where, and impact of their activities. Financial data today is available to almost anyone with a bank account and internet connection. Spending sensing is an easy-to-use and truly invisible method for sensing many aspects of human activity.

Each of our applications is intended to help individuals learn more about their habits and lives. They are more sophisticated versions of the line graphs and pie charts we see on financial management sites today. In the future we hope to explore how financial data can be combined with other sensed data. For example, the Errand Planner could leverage real-time location and machine learning to suggest new errands as people are doing them. These are just a few of the many applications that use financial data to help people lead more productive lives.

**ACKNOWLEDGMENTS**

removed for blind reviewing

**REFERENCES**