Reflections of Everyday Activities in Spending Data

Julia Schwarz[1], Jennifer Mankoff[2], and H. Scott Matthews[2]

[1]Dept. of CS, Univ. Wash.
Seattle, WA 98195-2350, USA
julenka@cs.washington.edu

[2]HCII, Carnegie Mellon
5000 Forbes Ave, Pittsburgh, PA 15213, USA
{jmankoff, hsm}@cmu.edu

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 limitations of sensing using financial data and possible solutions.

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 proposed the Memex, a device that would record 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].

Life-logging technologies currently require both a significant investment in infrastructure and/or access to unusual hardware. We propose to change this by leveraging a ubiquitously available, inexpensive source of data: financial transactions. In today's world, many things we do have some sort of financial impact. Financial data may show a warped and incomplete record of our activities, but it is nonetheless plentiful, more complete than many other digital sources of activity information, and readily available.

Personal financial data has recently become easily available in a consistent form through online banking, credit cards, and multi-account personal finance managers such as Wesabe.com and Microsoft Money. Financial transactions typically have a user or system-provided tag, a date, an amount, and a merchant associated with them. Our work leverages the fact that tagging of financial data is partly or fully automated by today's tools. Users are motivated to tag because it enables many powerful features in existing applications. The meta-data encoded in tags can be combined with other transaction information to extract 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. To implement them, we used a publicly available API of one of these financial managers (Wesabe) as our 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 converts the dollar value of certain tagged transactions into CO_2 emissions and displays environmental impact over time.
3. Personal Bubble combines merchant location information with financial information to create a visualization that shows people where they go and how to reduce travel distance when completing daily errands.

Our applications can sense the impacts of a wide range of activities, including arbitrary user-defined impacts. This may include social activities among friends, travel, home improvement, and entertainment. Different activities are reflected at different fidelities (*e.g.*, exercise is reflected over long periods of time by certain purchases such as apparel and club memberships, whereas transportation by auto results in frequent gas and possibly toll expenditures). Nevertheless, there are some limitations of sensing using financial data. We discuss these limitations and possible solutions at the end of the paper.

BACKGROUND

The analysis of personal financial data is a well-examined field. Personal finance managers automatically download transactions from multiple financial accounts and associate category tags with each transaction to better support reflection activities. While these tags may be aggregated across users to display consumer trends [1], most

CHI 2009, April 4–9, 2009, Boston, Massachusetts, USA.

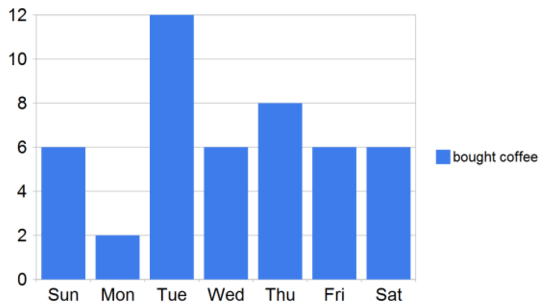

Figure 1: The author never drinks coffee on Mondays.

applications show an individual a pie-chart or other graph of her top spending categories.

Data mining can be used to create purchasing models for individuals. 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 the information about human activities encoded in financial data. The remainder of this paper provides examples 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.

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, and the Google Visualization API.

After a user loads data from Wesabe, he sees a bar chart of tagged transactions binned by weekday. The user can explore by selecting any tag or set of tags and viewing a bar chart of that subset of the data (Figure 1). The user can view the same transactions by date in a timeline view.

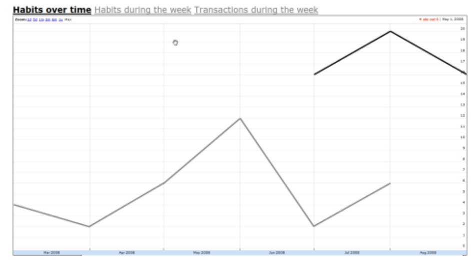

Figure 2: Eating out (grey) vs. eating in (black) over time.

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: Our validation is informal, but demonstrates that Snoopy can be used to explore personally relevant activities with real data. Using about six months of her own data, one author tracked a planned reduction in coffee drinking (by tracking dollars spent with the tag **`'coffee'`**). She learned that she rarely 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, which reflects a change in her location and lifestyle.

Discussion: Snoopy illustrates that it is possible to define habits in terms of financial transactions. Financial data can be used to show people when they do things, and how their habits change over time. Our validation illustrates the value of supporting end-user programming: the best way to define a habit is known by the person engaged in it. Snoopy could be improved by the addition of more sophisticated end-user programming techniques (*e.g.*, [10]).

CO2GREEN: IMPACT ON THE ENVIRONMENT

While Snoopy demonstrates how financial data can be used to track activities, CO2Green tracks the *impact* of activities.

Today, many online carbon calculators can calculate a person's emissions from self-reported information such as dollars spent on gas and kilowatt hours of electricity used. Different calculators vary in their level of detail, and there is a lack of consistency across calculators [9]. 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. Also, a survey cannot easily be used to calculate an individual's monthly or weekly carbon emissions. Since there is a strong connection between dollars spent and environmental impact [7], financial data provides an easy way to track carbon emissions over time.

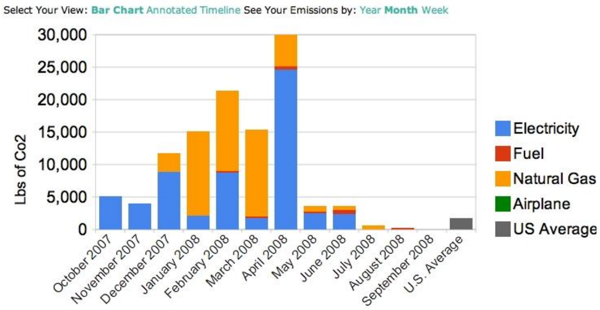

Figure 3: Monthly view of a six person household. The footprint rises in the winter, (the home's walls are not insulated) and falls in the summer (due to the shade of several large trees). This household does not drive or fly much.

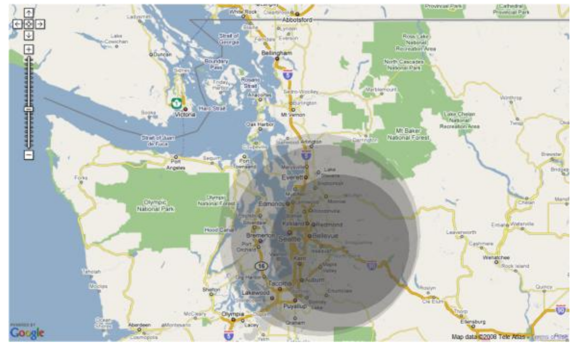

Figure 4: The Personal Bubble draws a circle around merchants you visit most frequently and recently. Darker bubbles represent more recent visits. The first author's bubble.

Application Description: Co2Green calculates the carbon emissions for every relevant transaction, and then displays these emissions over time. We used national average data from the year 2008 to convert spending dollars for car fuel, natural gas and electricity into carbon emissions [5, 8, 12]. 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 her data and specifies which tags indicate purchases of auto fuel, natural gas, electricity and airplane flights, she sees a bar graph of the calculated emissions over the last 12 months, grouped by category (fuel, natural gas, airplane, electricity) (Figure 3). Unlike standard calculators, Co2Green shows how emissions change by week, month or year. This may help show the impact of season, temperature, and circumstance on emissions.

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 age was between 31 and 40 and their median income was 20-30k. All were employed full time. Users, who had all used Wesabe before, reportedly spent about 5 minutes modifying the tags in their data in preparation for using Co2Green.

Users liked the monthly breakdown of emissions provided by Co2Green and they rated it as equal to or slightly lower than Zerofootprint on all usability measures. However, most users did not feel that Co2Green was accurately representing their footprint. 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.

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: VISUALIZING WHERE

Financial sensing also provides a low-cost method for determining *where* people are. 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.

The personal bubble visualizes *where* individuals spend most of their money (Figure 4). It also suggests how errands could be combined by finding the shortest travel route for completing multiple purchases in one day.

Application Description: After a user loads his financial data he sees his transactions grouped by day. He can then indicate the start and end address for a given day (usually his home). The application displays the route with the shortest travel distance on a map. 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 2 or more purchases. The largest number of purchases users made in one day was 8, and the average number of errands in a multi-errand run was 2.25. Our application calculated an average distance reduction of 16% between the optimal route and the least efficient route, for the first author (averaged across days).

Discussion: The Personal Bubble illustrates that spending information can be combined with other information (in this case the location of merchants) to create more useful applications. This idea could be expanded. For example, we would like to create an application that prompts users in real time to visit nearby shops that they may be interested in (based on places which they've shopped in recently). Other data combinations could further enhance context-aware applications. For example, a ride share application could leverage financial data to connect individuals who drive past each other's home and shop at the same place or enjoy the same entertainment locations. A reminder application could automatically determine when to send the user a grocery list based on when he usually shops

OPPORTUNITIES AND PROBLEMS

The applications just presented demonstrate the promise of financial data for enabling users to track activities and their impact over time. At the same time, financial data suffers from certain systematic inaccuracies, privacy limitations, and other constraints. Here we discuss some problems and possible solutions.

Multiple family members may share a financial account: In the case of a credit card, GPS data could be used to infer which family member actually made a purchase.

Cash transactions are not tracked: While it is becoming increasingly rare to use cash, this could be a real issue for some users. One possible solution would be to create a "smart wallet" which detects when users take our cash and prompts them to record what they bought.

Financial data is not available in real-time: Merchants often wait days before posting credit card purchases. The "smart wallet" mentioned before could potentially help.

Tagging requires user effort: As stated in the introduction, today's financial programs have developed sophisticated semi-automated approaches to tagging such as leveraging user history and databases of known merchants and common tags. These tags are readily available through tools such as Wesabe's API.

Location information is crucial and unavailable: Location is essentially another type of tag, and can benefit from similar techniques. In the future, financial data can be made more powerful by combining it with other types of sensors (such as GPS).

Financial information is personal and private: Privacy and security are an important concern in any application that senses human activity. The applications described here are intended for personal use, and should be treated with the same level of security as credit card and bank statements. A full analysis comparing our approach to the benefits and risks of corporate analysis of financial data and to other sensing methods is planned for future work.

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.

Our applications demonstrate the value of an easy-to-use and truly invisible method for sensing many aspects of human activity. In the future we hope to explore how financial data can be combined with other sensed data. We also plan to leverage machine learning to infer characteristics about activities by using financial data. This would enable the creation of a more sophisticated version of Snoopy that could take examples as input and use them to *learn* rules.

We also plan to examine how sharing information about personal activity with others can be used to motivate people. One example would be to motivate people to reduce their personal bubble by comparing their personal bubbles to others in similar locations.

ACKNOWLEDGMENTS

This work was supported by the CRA Distributed Mentorship Project, Zerofootprint.net, Intel, Google, IBM, and NSF IIS-0803733, IIS-0823723, IIS-0745885, and EEEC-540865.

REFERENCES

1. An Average Consumer's Spending. http://www.nytimes.com/interactive/2008/05/03/business/20080403_SPENDING_GRAPHIC.html.
2. Bush, V. (1945) As we may think. *The Atlantic Monthly*, 176(1):101-108.
3. Cadez, I. V., Smyth, P. & Mannila, H. (2001) Probabilistic modeling of transaction data with applications to profiling, visualization, and prediction. *ACM SIGKDD*'01, 37-46.
4. Gemmel, J., Williams, L., Wood, K., Lueder, R. & Bell, G. (2004) Passive capture and ensuing issues for personal lifetime stories. *CARPE*'04, 48-55.
5. Gasoline and Diesel Fuel Update. http://tonto.eia.doe.gov/oog/info/ gdu/gasdiesel.asp.
6. Lee, M. & Dey, A. (2008) Lifelogging memory appliance for people with episodic memory impairment. *Ubicomp'08*, 45-53.
7. Matthews, S & Lave, L. (2000) Applications of environmental valuation for determining externality costs. *Environ. Sci. Technol.*, 1390-1395.
8. Natural Gas Prices (2008) http://tonto.eia.doe.gov/dnav/ng/ ng_pri_sum_dcu_nus_m.htm.
9. Padgett, JP. , Steinemann, A.C., Clarke, J.H. & Vandenbergh, M.P. (2008) A Comparison of Carbon Calculators. *Environmental Impact Assessment Review*, 106 – 115.
10. Pane, JF & Myers, BA. (2000) Tabular and Textual Methods for Selecting Objects from a Group. *VL'00*, 157.
11. Using Financial Data to Extract Information About Social Networks. http://www.freepatentsonline.com/y2008/0082349.html.
12. U.S. Electric Utility Sales, Revenue and Average Retail Price of Electricity. http://www.eia.doe.gov/cneaf/electricity/page/at_a_glance/sales_tabs.html.