Using Low-Cost Sensing to Support Nutritional Awareness

Jennifer Mankoff, Gary Hsieh, Ho Chak Hung, Sharon Lee, Elizabeth Nitao

Electrical Engineering and Computer Science Department
University of California at Berkeley
Berkeley, CA 94720-1770
nutrition@cs.berkeley.edu

Abstract. Nutrition has a big impact on health, including major diseases such as heart disease, osteoporosis, and cancer. This paper presents an application designed to help people keep track of the nutrional content of foods they have eaten. Our work uses shopping receipts to generate suggestions about healthier food items that could help to supplement missing nutrients. We present our system design: a capture and access application that, based on shopping receipt data, provides access to ambiguous suggestions for more nutritious purchases. We also report results from one formative user study suggesting that receipts may provide enough information to extend our work by also estimating what people are actually eating, as opposed to simply what they are purchasing.

1 Introduction

Nutrition has a big impact on health, including major diseases such as heart disease, osteoporosis, and cancer [10]. Although awareness of nutrition varies in different parts of the world, many people do not know exactly how many servings of fruits, grains, vegetables, and fats they are eating [13], much less which nutrients are missing in their diet. For instance, people are often not aware that dairy foods vary widely in their effect on calcium excretion [17]. Even for those who are aware that change is needed, it is often perceived as complicated (7 out of 10 Americans believe this, for example [12]). This paper presents an application designed to help people keep track of the nutrient contents of food they have eaten.

Our goal is to support awareness of nutrition and suggest potential dietary changes. From a research perspective, this is difficult for several reasons common to ubiquitous computing applications but not explored in capture and access settings in the past. Like past capture and access applications [1], the interfaces for data collection and access in our system aim to be unobtrusive but available at appropriate times. However, our application involves significant inferencing, with large amounts of resulting ambiguity (an issue more common to other types of Ubiquitous Computing applications such as context-aware applications). Because of our emphasis on low-cost sensing, accuracy will always be limited. More accurate data would be prohibitively expensive to gather, in terms of either user time, user privacy, or number of devices and sensors. Ubicomp environments, in practice, have not used a large variety of sensors [5], and we are not attempting to address this issue. Instead, we give the user information about why we are suggesting each change, an approach common, for example, to recommender systems [16]. This transparency allows her to make an informed decision as to whether to follow the suggestion.

In our system, shown in Figure 1, data collection can be performed after a shopping trip or when bills are being sorted at the end of the week and involves a single swipe of a receipt with a handheld scanner. Our access interface is a shopping list, printed at the user's request and annotated with the reasons behind suggestions in order to help the user identify potential inaccuracies. This portable piece of paper provides suggestions at the most pertinent moment: when the user is making purchasing decisions.

Our shopping suggestions were inspired in part by the Nutrition through the Lifetime© project [19]. Shoppers stepped through an education program about healthy eating that suggested alternate purchasing behavior at multimedia kiosks in a supermarket, which also gave them weekly coupons. The results were evaluated by examining what shoppers purchased. This data, which was not used by the system, was recorded using time-consuming manual data entry. In the end, the supermarket coupons produced by the system appeared to be, by far, the most persuasive element.

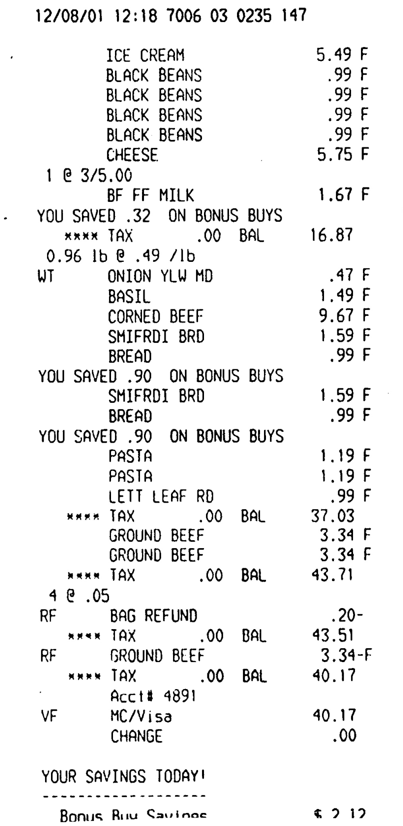
Computer scientists, nutritionists, and government agencies such as the USDA have been trying to better support nutrition since the 1960s [2,9,4,19,3]. However, none of them have overcome the cumbersome need to enter by hand everything one eats, a task which can take a minimum of 15-20 minutes per day [3].

This paper begins with a description of our user interface (a printed shopping list). The underlying technology used to generate this list (optical character recognition combined with a database inferencing system) is described next. We end with a description of some user studies that we have conducted and plan to do to address some of the more difficult problems raised in this application, including ambiguity, persuasion, and the provision of more intelligent suggestions.

2 User Interface

We chose to present suggestions to the user in the form of a printed shopping list that can be takend to the store. This gives the user information at the time when he can most use it: when making shopping decisions.

The shopping list, shown in Figure 1(b), is based on information from the most recent shopping trip (the receipt in Figure 1(a)), and suggests foods that are high in nutrients deficient from those foods just purchased. The first column shows suggested purchases, followed by the price per ounce, if known. Next is the main reason for the suggestion (usually a nutrient), followed by the previous purchase, for which the suggested food item is a healthier substitute. Finally, there are three checkboxes: "*bought*," "*helpful*" and "*not helpful.*" Although currently not implemented, the user will eventually be able to provide feedback to



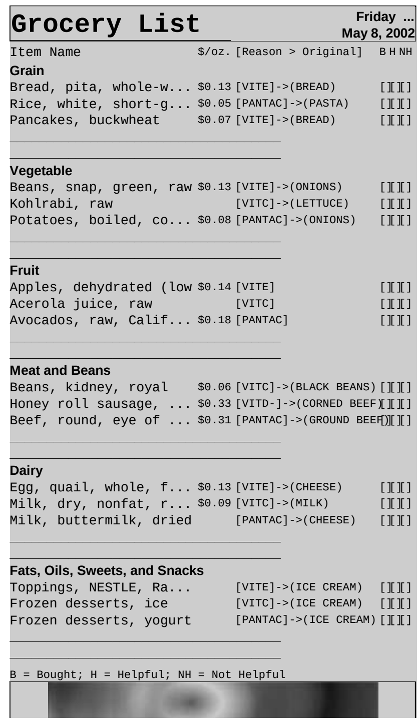
12/08/01 12:18 7006 03 0235 147

ICE CREAM 5.49 F
BLACK BEANS .99 F
BLACK BEANS .99 F
BLACK BEANS .99 F
BLACK BEANS .99 F
CHEESE 5.75 F
1 @ 3/5.00
BF FF MILK 1.67 F
YOU SAVED .32 ON BONUS BUYS
**** TAX .00 BAL 16.87
0.96 lb @ .49 /lb
WT ONION YLW MD .47 F
BASIL 1.49 F
CORNED BEEF 9.67 F
SMIFRDI BRD 1.59 F
BREAD .99 F
YOU SAVED .90 ON BONUS BUYS
SMIFRDI BRD 1.59 F
BREAD .99 F
YOU SAVED .90 ON BONUS BUYS
PASTA 1.19 F
PASTA 1.19 F
LETT LEAF RD .99 F
**** TAX .00 BAL 37.03
GROUND BEEF 3.34 F
GROUND BEEF 3.34 F
**** TAX .00 BAL 43.71
4 @ .05
RF BAG REFUND .20-
**** TAX .00 BAL 43.51
RF GROUND BEEF 3.34-F
**** TAX .00 BAL 40.17
Acct# 4891
VF MC/Visa 40.17
CHANGE .00

YOUR SAVINGS TODAY!

Bonus Buy Savings $ 2.12

(a)



Grocery List
Friday ...
May 8, 2002

Item Name	$/oz.	[Reason > Original]	B H NH
Grain			
Bread, pita, whole-w...	$0.13	[VITE]->(BREAD)	[][][]
Rice, white, short-g...	$0.05	[PANTAC]->(PASTA)	[][][]
Pancakes, buckwheat	$0.07	[VITE]->(BREAD)	[][][]
Vegetable			
Beans, snap, green, raw	$0.13	[VITE]->(ONIONS)	[][][]
Kohlrabi, raw		[VITC]->(LETTUCE)	[][][]
Potatoes, boiled, co...	$0.08	[PANTAC]->(ONIONS)	[][][]
Fruit			
Apples, dehydrated (low	$0.14	[VITE]	[][][]
Acerola juice, raw		[VITC]	[][][]
Avocados, raw, Calif...	$0.18	[PANTAC]	[][][]
Meat and Beans			
Beans, kidney, royal	$0.06	[VITC]->(BLACK BEANS)	[][][]
Honey roll sausage, ...	$0.33	[VITD-]->(CORNED BEEF)	[][][]
Beef, round, eye of ...	$0.31	[PANTAC]->(GROUND BEEF)	[][][]
Dairy			
Egg, quail, whole, f...	$0.13	[VITE]->(CHEESE)	[][][]
Milk, dry, nonfat, r...	$0.09	[VITC]->(MILK)	[][][]
Milk, buttermilk, dried		[PANTAC]->(CHEESE)	[][][]
Fats, Oils, Sweets, and Snacks			
Toppings, NESTLE, Ra...		[VITE]->(ICE CREAM)	[][][]
Frozen desserts, ice		[VITC]->(ICE CREAM)	[][][]
Frozen desserts, yogurt		[PANTAC]->(ICE CREAM)	[][][]

B = Bought; H = Helpful; NH = Not Helpful

(b)

Fig. 1. (a) A receipt from a local grocery store. **(b)** A shopping list generated by our system, based on that receipt.

the system about its suggestions by scanning in the shopping list. There is also space on the printout for the shopper to write in additional items.

The inferencing system used to generate these suggestions is described below. In practice, like any recognizer, our inferencing system will make mistakes. Interfaces designed for ambiguity may use several strategies to resolve it. Although there are many advantages to the use of paper, one limitation is the difficulty of using standard mediation techniques for resolving ambiguity [14]. For example, an interactive system may show the user a menu of choices, underline a word that may be incorrect, or take other steps to suggest to the user that an error may have occurred. We focused on transparent disclosure as a way of mitigating this [16]. By showing the shopper what the system believes he has bought, and why the alternative is believed to be better, we give the shopper the power of an informed veto. Additionally, we ensure that our suggestions *always* contain less fat and more nutrients overall than what was previously purchased (see the inferencing section below for how this is done). This means that even if we are not correctly addressing a specific nutritional deficiency, our suggestions will still be healthier overall than what was previously purchased.

Even if we were to make perfect suggestions, our users may not follow them. We use a shopping list as an interface in part because suggestions are more likely to be followed if they are made at pertinent times. We also examined other approaches to persuasion through a survey of food advertisements. A total of 30 food / health related advertisements were gathered from prime-time television over the course of two weeks. We found that advertisers used the following techniques to encourage shoppers to buy their food:

- Mouth-watering images to grab the attention of the viewers.
- Dialogs and skits to convey images of authenticity, deliciousness, cleanliness, quickness, and nutritiousness.
- Humor, music and sound to improve the ability to recall and associate the advertisement with a food.

Another strong source of persuasion is coupons [19]. Work similar to ours found that users' main motivation to purchase suggested items was not from the nutritional benefits these items would give them but because of coupons for these items. Interestingly, like our shopping list, coupons are right there when people are making purchasing decisions, an effect the authors did not measure directly. We hope that this will give us a similar advantage. Lastly, liking a food is critical to maintenance and change of dietary patterns [6].

3 Technology

Our system consists of three major stages. First, receipts are scanned in and passed through an optical character recognition (OCR) program. Second, data from this program is passed to a database system which records historical information and also stores important nutritional information about foods and nutrients. We expect users to scan their receipts on a semi-regular basis, perhaps at the same time they are doing their bills. The third component of the system is an inferencing system that estimates what the user eats on average per week, and compares this to recommended nutrient consumption. This then feeds in to the user interface described above.

3.1 Optical Character Recognition

A handheld HP Capshare is used to scan in each receipt which then is passed to a custom-built OCR system. The results are parsed into a list of foods, quantities, and prices. The OCR system and parser recognize 90% of characters correctly.

However, it is more informative to consider the number of food items on the receipt (such as "BLACK BEANS" "ICE CREAM" and so on) recognized correctly. We trained the system on 233 food items, and tested it on a further 20 items. The results of the testing were 80% accurate.

In addition to possibly being erroneous, the items listed on receipts are often abbreviated ("SMIFRDI BRD") or incomplete, so the next step is to match them against food names from our database. This is done using a combination of regular expressions and macros. When a food item does not match the food names exactly (because of a misrecognized character, for example), we look for the next closest match, where closeness is measured in number of characters substituted. The final data sent to the database is food names that match those in the database. The recognition accuracy at this stage is near perfect.

3.2 Database

We use a MySQL database to store profiles of our users, general data about nutrition, and historical information about user shopping habits. Data about the nutritional contents of foods comes from the USDA Nutrient Database for Standard Reference, Release 14 [15]. Data about the recommended daily allowances and dietary reference intakes (DRI) come from the USDA food and nutrition information center [7,8]. Other data comes from Nutrition Throughout the Life Cycle [20], including information about modifications to DRI for particular groups of people (based on age, gender, and other factors). This, plus a personal profile of age, weight, and specific health requirements, is used to calculate a personal DRI for each user. In the example in Figure 1, we assumed the shopper was a 19-30 year old woman, not pregnant or lactating.

The database also supports several functions. In addition to storing receipt data, the history can be queried for the amount of consumption of a specific nutrient. The query takes a time period and nutrient as input, and returns purchased foods that contributed significantly to that nutrient in that time period. The database can also be queried for a separate list of food items (not based on purchasing history) that provide significant amounts of a certain nutrient, and a list of the benefits of a specific nutrient.

3.3 Inferencing System

The inferencing system takes in a shopping list and outputs suggestions for healthier purchases. The food list produced by the OCR program may match multiple brands of that food stored in the database. The nutrient content of possible matches is averaged (for example, in the case of "SMIFRDI BRD", the nutrient contents of wheat and white bread will be averaged).

When it is time to give the user feedback, the nutrients/100 grams for the most recent purchases is compared to recommended daily values, and differences greater than ten percent are noted. In order to keep the number of alternatives small enough to handle, we only select a food if it has at least twenty percent more of the missing nutrient than the original for which it is a substitute. We also make sure it is in the same food group as the original. Finally, we weigh alternatives with less fat and more overall daily nutrients more highly. The top alternative for each original purchase is displayed on the final shopping list.

4 Moving beyond simple suggestions

The shopping suggestions we describe above are based on purchases, as opposed to actual food consumption. Extending this to consumption could improve the accuracy of our system but only if we can successfully infer what is consumed, despite potentially large families, eating out, and food waste. We present a survey that suggests the feasibility of this approach.

We surveyed shoppers at a local grocery store chain to determine the relationship between consumption and purchasing. We asked each shopper to fill out a short form detailing their consumption of calcium-related foods (foods that contain calcium or affect calcium absorption). We also requested their receipts, allowing us to compare their purchases to their reported consumption.

We focused on calcium in this survey for several reasons. First, most people know what foods are good sources of calcium, and thus can give us information to compare against what we infer from their receipts. Studies have shown that self-reported consumption is a reliable way to estimate eating habits [18], although more recent work may contradict this [13]. Second, many people have allergies to calcium-rich products such as milk, or age-related reasons to consume calcium, and thus intentionally buy foods containing that mineral. Third, a small number of foods contain high amounts of calcium, allowing us to create a small, precise and targeted survey.

Our participants included 57 people, from a range of economic and cultural backgrounds. Ages varied from 18 to 77, and the male/female split was approximately 40%/60%. The Body Mass Index (BMI) of our participants, a measure of health, varied from 15 to 43, with a standard deviation of 4.43, where 20-25 is a normal, healthy value, while anything over 30 is obese.

We also calculated the number of times a participant said they ate a food for health reasons (*health choices*), and correlated this with BMI and gender. Although the average BMI hovered around 24, the range of BMI was much greater for people with fewer health choices. The standard deviation is 5.99 for people with fewer than seven health choices, drops to 3.92 for people with between seven and eleven health choices, and down to 1.79 for people with between 12 and 15 health choices. People in the the last category are all healthy, with BMIs in the 20-25 range. In other words, the more aware a participant was of nutrition, the more likely he or she was to be healthy.

The correlation between food purchases and eating habits was 0.99–when people reported eating more of a food item, we found that they purchased proportionally more of it as well. We estimated consumption as the number of servings purchased (based on price and number) times the reported shopping frequency, divided by the number of family members. The result was off by about a factor of two.

We attribute the discrepancy to a lack of certain key information. We did not have accurate data on how well food price correlates with food amount. Also, we only had data on a single shopping cycle for each survey participant, so it was not possible to account for the fact that a participant might only purchase some things every second or third shopping trip. Finally, we did not know much about the age or eating habits of family members, and we over compensated for this by estimating that every family member ate the same amount of food.

In order to improve our algorithm for estimating consumption, we are embarking on a longer term survey in which we track the reported and purchased consumption habits of specific individuals over the course of a three-month period, using receipt data and a variant of the food frequency questionnaire [11]. Our participants will share food consumption data with us on a weekly basis, and give us all of their receipts. This data will allow us to improve our inferencing, and to experiment with different data aggregation periods to determine what is optimal for each food group. It will also allow us to understand the potential impact of household size and eating-out on the accuracy of our inferencing system. Depending on the success of our results, and user input, we may decide to supplement our inferences with occasional self-reports.

5 Summary and Future Work

We have presented a system for helping consumers make more nutritious choices when shopping for food. This paper focuses on the technology and inferencing necessary to generate helpful suggestions. The system currently suggests alternatives to purchases. Although this usually produces more nutritious alternatives, lack of data about what people prefer to eat still leads to bad suggestions (for example substituting “Quail Eggs” for “Cheese” in Figure 1). We plan to extend our algorithm to use historically based estimates of what people are actually eating to further improve the relevance of the suggestions. We are also in the process of conducting a longer term study to improve our ability to estimate portion sizes, a crucial component of our inferencing algorithm. On the interface side, we plan to explore the use of ambient displays, mouth-watering images on our printed shopping lists, and other persuasive mechanisms.

From a ubiquitous computing perspective, this work is important because it involves high-level inferencing about sensed data, and integrates the ambiguous results into a capture and access system. We plan to extend this work both by improving the inferencing, and investigating the interface implications of ambiguity in a capture and access setting.

6 Acknowledgments

We thank Cindy Shieh for her help with the nutritional side of this work. This work was supported in part by Lucent. Thanks to Bren Ahearn and Shiela Humphries for their support of the undergraduates in this undergraduate-only research project.

References

1. G. D. Abowd and E. D. Mynatt. Charting past, present, and future research in ubiquitous computing. *ACM TOCHI*, 7(1):29–58, 2000.
2. J. L. Balintfy. Menu planning by computer. *Communications of the ACM*, 7(4):255–259, 1964.
3. J. Beidler *et al.* The PNA project. In *Proceedings of the sixth annual CCSC northeastern conference on The journal of computing in small colleges*, pp. 276–284. The Consortium for Computing in Small Colleges, 2001.
4. CNPP. Interactive healthy eating index. Powered by Commerce One `http://147.208.9.133/`.
5. A. K. Dey, G. D. Abowd, and D. Salber. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human Computer Interaction*, 16(2-4):97–166, 2001.
6. A. Eertmans, F. Baeyens, and O. V. den Bergh. Food likes and their relative importance in human eating behavior: Review and preliminary suggestions for health promotion. *Health Education Research*, 16(4):443–456, 2001.
7. Food and Nutrition Information Center. Recommended daily allowances, 1989.
8. Food and Nutrition Information Center. Dietary reference intakes, 2001.
9. A. Garrison, S. J. Mountford, and G. Thomas. Designing computers with people in mind. *interactions*, 1(2):61–72, 1994.
10. S. Hankinson *et al.*, editors. *Healthy Women, Healthy Lives: A Guide to Preventing Disease*. A Harvard Medical School book. Simon & Schuster Inc., 2001.
11. F. Hu *et al.* Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. *American Journal of Clinical Nutrition*, 69(2):243–249, February 1999.
12. INSIGHT 19. Beliefs and attitudes of americans toward their diet. USDA Center for Nutrition Policy and Promotion, June 2000.
13. INSIGHT 20. Consumption of food group servings: People's perceptions vs. reality. USDA Center for Nutrition Policy and Promotion, October 2000.
14. J. Mankoff, S. E. Hudson, and G. D. Abowd. Interaction techniques for ambiguity resolution in recognition-based interfaces. In *Proceedings of the ACM UIST 2000*, pp. 11–20. ACM Press, November 2000.
15. Nutrient Data Laboratory. Usda nutrient database for standard reference, release 14. Nutrient Data Laboratory Web Page `www.nal.usda.gov/fnic/foodcomp`.
16. K. Swearingen and R. Sinha. Interaction design for recommender systems. In *DIS 2002*. ACM, 2002.
17. R. Weinsier and C. Krumdieck. Dairy foods and bone health: examination of the evidence. *American Journal of Clinical Nutrition*, 72(3):681–689, September 2000.
18. W. Willet. Nutritional epidemiology: issues and challenges. *International Journal of Epidemiology*, 16(2):312–317, June 1987.
19. R. A. Winett, *et al.* Nutrition for a lifetime system: A multimedia system for altering food supermarket shoppers? purchases to meet nutritional guidelines. *Computers in Human Behavior*, 13(3):371–392, August 1997.
20. B. S. Worthington-Roberts and S. R. Williams. *Nutrition Throughout the Life Cycle*. McGraw Hill, 2000.