2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2013)
September 8-12, 2013, Zurich, Switzerland

Placer: Semantic Place Labels from Diary Data

John Krumm and Dany Rouhana
Microsoft Research
Microsoft Corporation
Redmond, WA USA
{jckrumm|danyr}|@microsoft.com

ABSTRACT
Semantic place labels are labels like "home", "work", and "school" given to geographic locations where a person spends time. Such labels are important both for giving understandable location information to people and for automatically inferring activities. Deployed products often compute semantic labels with heuristics, which are difficult to program reliably. In this paper, we develop Placer, an algorithm to infer semantic places labels. It uses data from two large, government diary studies to create a principled algorithm for labeling places based on machine learning. Our labeling reduces to a classification problem, where we classify locations into different label categories based on individual demographics, the timing of visits, and nearby businesses. Using these government studies gives us an unprecedented amount of training and test data. For instance, one of our experiments used training data from 87,600 place visits (from 10,372 distinct people) evaluated on 1,135,053 visits (from 124,517 distinct people). We show labeling accuracy for a number of experiments, including one that gives a 14 percentage point increase in accuracy when labeling is a function of nearby businesses in addition to demographic and time features. We also test on GPS data from 28 subjects.

Author Keywords
Semantic place labels, location, ATUS, PSRC.

ACM Classification Keywords
I.5.4. Pattern Recognition: Applications.

INTRODUCTION
Semantic place labeling is the process of giving a meaningful name to a location. For example, the process might give the label "home" to the geographic location where a person lives, "work" to their workplace (Figure 1), "school" to school, and so on. This labeling is important when we want an application to refer to locations in an understandable way, such as, "Barney has arrived safely at school." Place labels

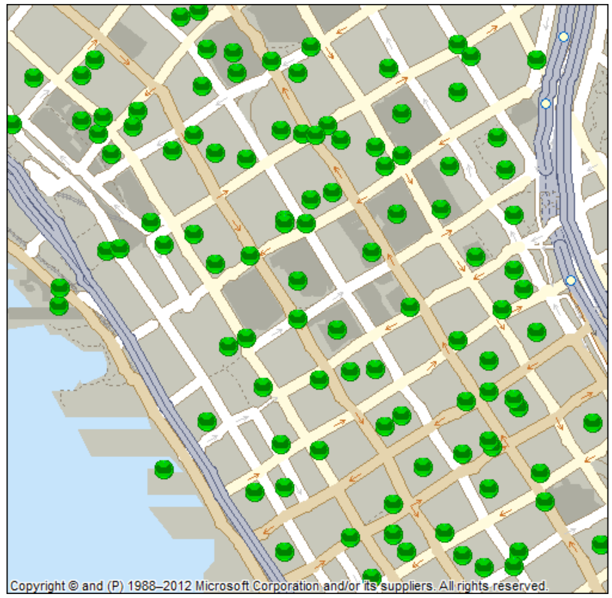

Figure 1: These are locations labeled as "work" in a diary study. Our goal is label them automatically based on features including the timing of visits, demographics of the visitors, and surrounding businesses.

have also been proposed for automatically updating a person's status on social networking sites, such as the CenseMe project [1], and automatically annotating check-ins, such as work by Ye *et al*. [2]. Friendly names for locations are much easier to understand than latitude/longitude or street addresses. Semantic location labels can also serve as input to automatic activity inference. For instance, sleep usually occurs at home rather than at work or school (although there are exceptions). In fact, this ability to infer an activity from a labeled location was shown by Partridge and Golle [3] using the American Time Use Survey [4], which is one of the two diary studies we use.

In this paper, we develop a classifier called Placer that identifies place labels based on the timing of visits to that place, nearby businesses, and simple demographics of the user. For example, if a student-age person arrives at a location near a school on a winter weekday at 8:00 a.m. and stays until 3:00 p.m., our classifier would say that place is the person's school. This is an example of a heuristic that could be written to label a place. Such heuristics have been tested in [5] for identifying a person's home and work place. For instance, a person's home was identified as where they spent the most time during the hours of midnight to 6 a.m. It would be cumbersome and error-prone to write such heuristics that cover all possible cases and places. Instead, modern approaches, including ours and [5], use machine learning to build a place classifier automatically.

Other approaches to place labeling include those that collect labels from other people. Loci [6] gives users a suggested list of labels based on labels left by other users. Similarly, CenceMe [1] automatically labels a user's place based on labels given to the place by friends. Our approach is different in that it attempts to automatically label places based on how an individual uses them and the surrounding businesses.

One of the feature types we use to infer a place label is nearby businesses and points of interest. Intuitively, a school at a place gives evidence that the place is either a person's school or workplace. Several researchers have shown how to characterize places by their nearby points of interest (POI) [7-9]. Our work goes a step beyond characterization and shows how to infer an actual label from the POI, as well as other features.

With straightforward machine learning, our Placer approach creates a mapping between features describing a visit to a place and the place's label. One of the earliest attempts to do this was the work by Liao *et al*. [10], who used features including the timing of visits and the presence of bus stops, restaurants, and grocery stores. Their main innovation was a hierarchical conditional random field (CRF) that aided inference accuracy by exploiting the temporal sequence of place visits, *e.g.* "work" often follows "home". Their system was tested on GPS traces of four people. Chen *et al*. [11] processed label sequences with a hidden Markov model rather than a CRF. Ye *et al*. [2] derived eight place label categories from Whrrl, a location-based social network. They used a support vector machine on features such as check-in frequency and time of day to label over 53,000 places from almost 6000 users. Based on our literature search, this is the largest test of place labeling to date. Nokia's Mobile Data Challenge (MDC) attracted several algorithms for inferring semantic place labels (MDC Task 1) [12]. The MDC supplied data sensed from mobile phones (time, battery state, accelerometer, Bluetooth, WiFi, call log, SMS log, and selected user profile) for 79 users with an average of 5 labeled locations each. The published solutions, all of which used a machine learning approach, had classification accuracies between 65% and 75% [5, 13-15] for the 10 types of places specified.

Our approach to semantic place labeling, called Placer, solves the labeling task as a classification problem, similar to the solutions above. We train and test our approach on two different diary survey conducted in the U.S. These large datasets, one of which includes latitude/longitude data, give us the opportunity to move beyond existing research in the following ways:

- Train/test on an order of magnitude more data than previously largest test.
- Include relatively low frequency visits to places such as banks, post offices, and churches.
- Include demographic features (age and gender) for the first time.
- Differential testing to assess the accuracy impact of including an extensive set of nearby business features.
- Testing across two different, large data sets to assess the impact of inevitable coding inconsistencies.
- Apply diary survey model to real GPS data.

We elaborate on these advantages subsequently, but next we describe the public datasets we used.

DIARY DATASETS

We used two publicly available location diaries for our training and testing. In both, subjects were asked to keep track of where they went for one or two days. The two studies are the American Time Use Survey (ATUS) [4] and the Puget Sound Regional Council Household Activity Survey (PSRC) [16]. These are both examples of diary surveys that exist in many countries. Partridge and Golle [3] give a good overview of these types of surveys and introduce some uses of them for understanding people's activities and places.

American Time Use Survey

The goal of ATUS [4] is to understand how Americans spend their time in different activities. The survey is conducted by the U.S. Census Bureau. Survey subjects are randomly chosen, and they must be at least 15 years old. The survey is conducted entirely by telephone, and subjects are asked to recall their previous day's activities and trips. For our purposes, the important questions are about the subject's age, gender, and when and where they went. Each row in our ATUS database table corresponds to one visit to a place by one of the subjects, and each visit includes the visit's date, start time, duration, and place category. For each visit, the subject chose one of the place categories shown in the first column of Table 1. We used ATUS data from 2003-2011, giving a total of 124,517 unique subjects who recorded 1,135,053 total visits, giving an average of 9.1 recorded visits per user. To the best of our knowledge, this is the largest dataset to be used for semantic place labeling to date, with the next largest being the 53,432 visits by 5892 subjects from the Whrrl data in [2].

Puget Sound Regional Council Household Activity Survey

The second diary study we used is the Puget Sound Regional Council (PSRC) Household Activity Survey [16]. The goal of this survey, conducted in 2006, was to discover travel trends for the U.S. Puget Sound region, which consists of four U.S. counties anchored by Seattle, Washington. The

ATUS Where	AUTS Where for Placer	Generic Where for Placer
Respondent's home or yard	Home	Other
Respondent's workplace	Work	Work
Someone else's home	Someone else's home	Other
Restaurant or bar	Restaurant or Bar	Restaurant or Bar
Place of worship	Place of Worship	Place of worship
Grocery store	Store for Shopping	Store for Shopping
Other store/mall	Store for Shopping	Store for Shopping
School	School	School
Outdoors away from home	Outdoors	Other
Library	Library	Other
Other place	Other	Other
Car, truck, or motorcycle (driver)	Transportation	Other
Car, truck, or motorcycle (passenger)	Transportation	Other
Walking	Transportation	Other
Bus	Transportation	Other
Subway/train	Transportation	Other
Bicycle	Transportation	Other
Boat/ferry	Transportation	Other
Taxi/limousine service	Transportation	Other
Airplane	Transportation	Other
Other mode of transportation	Transportation	Other
Bank	Bank	Other
Gym/health club	Gym	Other
Post Office	Post Office	Other
Unspecified place	IGNORED	IGNORED
Unspecified mode of transportation	Transportation	Other

Table 1: These are the place categories from ATUS and the translations we made for Placer.

PSRC Activities	PSRC Where for Placer	Generic Where for Placer
Home - Paid Work	Work	Work
Home - Other	Home	Home
Work	Work	Work
Attend Childcare	School	School
Attend School	School	School
Attend College	School	School
Eat Out	Restaurant or Bar	Restaurant or Bar
Personal Business	Personal Business	Other
Everyday Shopping	Store for Shopping	Store for Shopping
Major Shopping	Store for Shopping	Store for Shopping
Religious/Community	Religious/Community	Place of Worship
Social	Social	Other
Recreation - Participate	Recreation	Other
Recreation - Watch	Recreation	Other
Accompany Another Person	Accompany Another Person	Other
Pick-Up/Drop-Off Passsenger	Pick-Up/Drop-Off Passsenger	Other
Turn Around	Turn Around	Other

Table 2: These are the place categories from PSRC and the translations we made for Placer, analogous to the ATUS place categories in Table 1.

subjects of this survey filled out a travel diary, giving details of their trips for two consecutive days. For each trip, the subjects indicated the purpose of the trip from the list of choices shown in Table 2. From the trip data, we could easily derive each visit's date, start time, and duration, as in the ATUS data. The PSRC survey also came with the age and gender of each subject, matching the demographic data available from ATUS. By signing a confidentiality agreement with PSRC, we were also able to obtain extra data giving the geographic coordinates of each subject's visits. We explain later how we used this coordinate data to create a more accurate inference model of semantic place labels.

We used PSRC data on 87,600 trips taken by 10,372 different people. This PSRC survey was budgeted for one million U.S. dollars, which is more than any research institution would likely spend on the problem of semantic place labeling. Thus we are fortunate that this data is freely available and can be used for purposes beyond its original intention. In the next section we explain how we used both the ATUS and PSRC data to infer place labels with machine learning.

MACHINE LEARNING FOR SEMANTIC PLACE LABELING

Our goal is to compute a semantic place label from data on visits to the place. We might reasonably assume that visits to a school have different patterns than visits to a restaurant, especially temporally. In our baseline approach to this task, we use demographic (age and gender) features and temporal features of the visit to infer a place label. Specifically, from both ATUS and PSRC, our baseline features for each visit are the following, all scalars:

- Age of subject in integer years
- Gender of subject
- Arrival day of week
- Arrival time of day
- Visit midpoint time of day
- Departure time of day
- Duration of visit
- Holiday (binary)
- Season of year (0,1,2,3)

Some of the nine baseline features are redundant in that they can be computed from each other. This was intentional to make machine learning easier. Each feature vector came with a ground truth semantic place label from either ATUS or PSRC.

The mapping between the feature vector and place label is computed with a learned multiclass classifier in the form of a forest of boosted decision trees [17]. This learning process begins with learning a conventional decision tree. Using the classification results from this tree, a second tree is learned with increased importance given to the training samples that were misclassified by the first tree. More trees are added in this way to create the forest. Given a feature vector, the forest gives a probability for each class, and we take the highest probability class as the inference.

For our particular learner, the parameters were:

- Maximum branching factor: 20
- Minimum instances per leaf: 10
- Learning rate: 0.2
- Number of trees: 100

We felt some of the labels given by ATUS and PSRC were unnecessarily precise for many applications, such as PSRC's separate labels "Everyday shopping" and "Major shopping". We combined these into one label category called "Store for Shopping". Similarly, we combined ATUS's 11 different transportation labels into one "Transportation" category. We also ignored all ATUS records labeled with "Unspecified place". These changes are highlighted in the second columns of Table 1 (ATUS) and Table 2 (PSRC). With these changes, we had 14 different place labels for ATUS and 13 for PSRC.

ATUS Place Label Inferences

We tested our inferences on the ATUS data using 10-fold cross validation. The results, given as a confusion matrix, are

Inferred Label

True Label	Home	Work	Other's Home	Restaurant or Bar	Place of Worship	Store for Shopping	School	Outdoors	Library	Other Place	Transportation	Bank	Gym	Post Office	Recall
Home	0.92	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.01	0.92
Work	0.05	0.87	0.02	0.01	0.00	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.01	0.87
Other's Home	0.15	0.07	0.68	0.03	0.01	0.02	0.01	0.01	0.00	0.00	0.00	0.01	0.00	0.01	0.68
Restaurant or Bar	0.19	0.11	0.08	0.30	0.03	0.08	0.06	0.05	0.02	0.01	0.00	0.01	0.00	0.06	0.30
Place of Worship	0.25	0.03	0.04	0.05	0.33	0.05	0.12	0.04	0.03	0.00	0.00	0.01	0.00	0.05	0.33
Store for Shopping	0.16	0.06	0.02	0.07	0.03	0.40	0.05	0.10	0.06	0.00	0.00	0.01	0.01	0.04	0.40
School	0.19	0.05	0.01	0.03	0.05	0.04	0.44	0.04	0.07	0.02	0.00	0.00	0.01	0.04	0.44
Outdoors	0.17	0.05	0.02	0.02	0.02	0.09	0.05	0.45	0.02	0.04	0.00	0.01	0.00	0.05	0.45
Library	0.17	0.06	0.03	0.03	0.02	0.05	0.09	0.03	0.45	0.01	0.00	0.00	0.01	0.04	0.45
Other Place	0.20	0.08	0.05	0.04	0.02	0.03	0.05	0.07	0.03	0.36	0.00	0.00	0.01	0.06	0.36
Transportation	0.19	0.16	0.05	0.04	0.03	0.05	0.07	0.07	0.10	0.02	0.15	0.00	0.01	0.06	0.15
Bank	0.34	0.08	0.11	0.05	0.05	0.11	0.05	0.05	0.02	0.03	0.01	0.08	0.00	0.03	0.08
Gym	0.35	0.10	0.01	0.00	0.00	0.13	0.06	0.06	0.15	0.02	0.01	0.00	0.10	0.01	0.10
Post Office	0.26	0.11	0.07	0.08	0.05	0.05	0.09	0.08	0.05	0.03	0.00	0.00	0.00	0.12	0.12
Precision	0.85	0.81	0.67	0.38	0.38	0.38	0.41	0.40	0.41	0.47	0.54	0.26	0.26	0.27	

Table 3: This is the confusion matrix from a 10-fold cross validation with the ATUS data. The overall accuracy was 0.73. "Home" and "Work" stand out as the easiest-to-label places.

shown in Table 3. Here we see that "Home" and "Work" stand out as the most accurate inferences. "Other's Home" has fair accuracy at 0.68, while all other places are mistaken for the wrong place over half the time. But, since people spend much time at home and work, the overall accuracy for ATUS is 0.73. We computed this accuracy number by simply dividing the number of correct classifications by the total number of classification attempts, so it naturally accounts for the fact that there are many more visits to home and work than to other places.

PSRC Place Label Inferences

The PSRC data has a different set of place labels, so we cannot compare it directly with the ATUS data. The other major difference between the ATUS data and the PSRC data is that the PSRC data comes with geographic coordinates giving the location of each visit. This gave us the opportunity to add classification features pertaining to the characteristics of the actual place, in addition to the demographic and temporal features that we used for the ATUS data.

We used data from a local search engine to find businesses and points of interest (POI) near every visit in the PSRC data. Our intuition was that geographic features like this would help distinguish place types. For instance, the presence of a school is indicative of the place type. Instead of specific businesses (*e.g.* Starbucks) and POI (*e.g.* Benjamin Rush Elementary School), we instead looked at 15 types of geographic entities:

- Arts & Entertainment
- Automotive & Vehicles
- Business to Business
- Computers & Technology
- Education
- Food & Dining
- Government & Community
- Health & Beauty
- Home & Family
- Legal & Finance
- Professionals & Services
- Real Estate & Construction
- Shopping
- Sports & Recreation
- Travel

For each of these 15 types, we created 4 different features for classification:

- Count of each type with 50 meters
- Count of each type within 100 meters
- Count of each type within 200 meters
- Distance to nearest instance of each type

This created 60 additional features for classification to supplement the 9 baseline features we used for the ATUS data. We used the same machine learning model and parameters as for the ATUS data.

The results are shown in Table 4 as a confusion matrix. The overall accuracy is 0.74, which is essentially the same as the overall accuracy for the ATUS result. While both ATUS and PSRC do fairly well on "Home", "Work", and "School", the confusion matrix shows that PSRC is also doing fairly well on less frequently visited locations such as "Personal Business", "Shopping", and "Religious/Community".

We suspect this improved performance is due to the extra geographic features that PSRC lets us use. We tested this by rerunning our test on the PSRC data without the geographic features, using just the nine baseline demographic and temporal features we used for the ATUS data. Without the

Inferred Label

True Label	Home	Work	School	Eat Out	Personal Business	Store for Shopping	Religious/Community	Social	Recreation	Accompany Person	Pick-Up/Drop-Off	Turn Around	Recall
Home	0.91	0.02	0.01	0.00	0.01	0.01	0.01	0.00	0.00	0.01	0.01	0.00	0.91
Work	0.07	0.83	0.01	0.01	0.03	0.02	0.01	0.01	0.00	0.01	0.01	0.00	0.83
School	0.04	0.02	0.88	0.00	0.01	0.00	0.02	0.01	0.00	0.01	0.01	0.00	0.88
Eat Out	0.02	0.02	0.01	0.66	0.07	0.12	0.00	0.03	0.01	0.03	0.02	0.01	0.66
Personal Business	0.14	0.04	0.01	0.02	0.58	0.10	0.03	0.02	0.01	0.03	0.02	0.01	0.58
Store for Shopping	0.10	0.03	0.01	0.03	0.07	0.62	0.05	0.05	0.01	0.01	0.02	0.02	0.62
Religious/Community	0.14	0.02	0.02	0.00	0.04	0.06	0.57	0.09	0.00	0.02	0.02	0.02	0.57
Social	0.06	0.02	0.01	0.02	0.02	0.06	0.08	0.68	0.00	0.02	0.01	0.02	0.68
Recreation	0.26	0.06	0.02	0.03	0.06	0.06	0.02	0.01	0.34	0.06	0.07	0.01	0.34
Accompany Person	0.21	0.05	0.01	0.03	0.05	0.04	0.04	0.03	0.02	0.46	0.05	0.01	0.46
Pick-Up/Drop-Off	0.14	0.03	0.02	0.01	0.05	0.05	0.03	0.03	0.01	0.04	0.57	0.02	0.57
Turn Around	0.19	0.01	0.01	0.00	0.04	0.09	0.08	0.04	0.01	0.02	0.04	0.47	0.47
Precision	0.82	0.83	0.89	0.70	0.64	0.64	0.62	0.64	0.50	0.57	0.64	0.60	

Table 4: This is the confusion matrix from a 10-fold cross validation with the PSRC data that includes business features. The overall accuracy was 0.74. "Home", "Work", and "School" stand out as the easiest-to-label places.

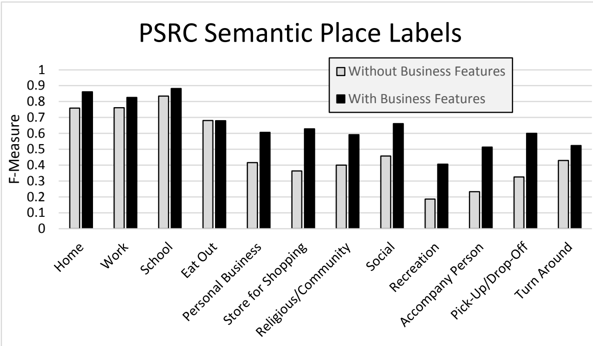

Figure 2: Adding business features improves the performance of place labeling for almost every place category in the PSRC data.

geographic features, the overall PSRC accuracy was 0.60, meaning that geographic features improved accuracy by 14 percentage points. The F-Measures for each place category are shown in Figure 2. (The F-Measure is the harmonic mean of precision and recall.) The addition of geographic features substantially improves the F-Measure for every place category except “Eat Out”, where it has essentially no effect.

While the geographic features may seem like a simple solution to labeling places like schools, stores, and restaurants, these places may instead be labeled by some users as a workplace, which can be distinguished by the baseline features pertaining to demographics and timing.

ATUS and PSRC on Common Place Categories

It is instructive to train a model based on the ATUS data and test on the PSRC data, and vice-versa. This helps reveal the effects of the inevitable coding inconsistencies between the two surveys. We know, for instance, that subjects in the ATUS survey filled their diaries from memory over the phone the day after their survey day, while PSRC subjects had a diary form to fill out themselves at any point during or after the survey. Such a cross test also gives us an idea of how effective it would be to use a learned ATUS or PSRC model on sensed location data, since diary data is likely cleaner than sensed data.

We can test across ATUS and PSRC if we match both their feature types and their place categories. Matching their feature types is easy, because we can just use the nine baseline demographic and temporal features that both surveys support. We cannot use geographic features, because ATUS does not come with coordinate data.

Matching place categories is slightly more complicated. We are forced to find commonalities between the two surveys’ categories based on the category names. We can confidently match ATUS’s “Respondent’s home or yard” to PSRC’s “Home – other” category. Other categories are just as clear, but there are exceptions. We took a fairly conservative approach: instead of guessing on ambiguous matches, we grouped uncertain matches into an “Other” category. Perhaps the most questionable match we made was ATUS’s “Place of worship” to PSRC’s “Religious/community”. The matches we chose are shown in the third columns of Table 1 (ATUS) and Table 2 (PSRC).

With matches between features and place categories, we can test and train on ATUS and PSRC. Figure 3 shows the results for the common place categories in terms of the F-measure as well as overall accuracy. Note that these tests are different than those in the previous sections, because here we used the common set of place categories instead of those that are specific to ATUS or PSRC.

For comparison, the first three tests show ATUS and PSRC trained and tested on themselves, not mixed with each other, with 10-fold cross validation. There are two versions of the purely PSRC test: one using just the nine demographic and temporal features and one using those plus the business features. In terms of overall accuracy between these three, the pure ATUS test has the best result, with PSRC with business features close behind. PSRC without business features lags behind these three non-mixed results in terms of overall accuracy, although it trades places with the others in terms of F-measures on the place categories. Looking at the categories, we see that PSRC with business features is most often the winner, supporting our earlier finding that geographic features are important. These features appear especially helpful on the less frequent categories of “Eat Out”, “Religious/Community” and “Shopping”.

The right-most columns of each cluster in Figure 3 show the results of training on ATUS and testing on PSRC, and vice-versa. This gives our most data-intensive tests, with 1,135,053 ATUS visits and 87,600 PSRC visits. We see that these cross tests often give worse results than the pure tests, possibly because of inconsistent coding between the two surveys. The category of “Religious/Community” has low F-measures for both mixed tests. This may be because ATUS’s “Place of Worship” category is possibly quite different from PSRC’s “Religious/Community” category. Regional differences may also play a role, as the PSRC data comes from one relatively small part of the U.S.

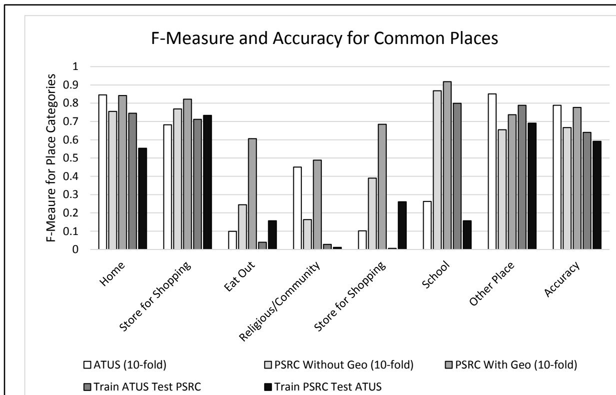

Figure 3: F-Measures and accuracies of place labeling within and between ATUS and PSRC.

TEST ON GPS DATA

Thus far we have tested Placer on ATUS and PSRC data. In this section we use our PSRC model to label home and work locations from GPS data of real users.

We began by logging GPS data from 28 volunteers (7 female) at our institution. Each subject borrowed a RoyalTek RBT-2300 GPS logger and placed it in their main vehicle, powered by the cigarette lighter. Most of the subjects were compensated with a US$ 30 cafeteria card, although a few agreed to participate without compensation. Our goal was to collect at least six weeks of data from each subject. In the end, we obtained data for an average of 47.6 days, varying from 27 to 60 days.

The loggers were set to record a time-stamped latitude/longitude pair every 10 seconds. Since the logger was powered by the vehicle's cigarette lighter, some loggers turned off when the vehicle was turned off, while others logged continuously.

Point Clustering

We performed simple processing to extract the locations of visits. This is a well-studied problem in the mobile computing research literature, including work by Marmasse and Schmandt [18], Ashbrook and Starner [19], and Kang *et al.* [20], Hightower *et al.* [21], and Caoy *et al*. [22]. We adopt a simple agglomerative clustering technique that starts by treating each individual GPS point as its own cluster. It then creates a new cluster by merging the two geographically nearest clusters into a new cluster and deleting the two constituent clusters. Merging continues like this until all the clusters are at least 100 meters apart.

Although we value the simplicity of this technique, we were forced to modify it in two ways. One modification was to account for the warm-up time for GPS. If the GPS logger turned off with the cigarette lighter, then it could take up to a minute to start logging on the next trip as the vehicle drives away from its last visit. Even if the GPS stays on, there can be a similar delay if the vehicle is parked where it loses its view of the GPS satellites, such as a parking garage. To account for these situations, we artificially repeated the most recent latitude/longitude point after each gap in logging that exceeded the sampling time of 10 seconds. This had the effect of inserting a GPS point close to the time and place of a departure after being parked.

The other modification was designed to ignore GPS points where the vehicle was moving. This helped avoid small clusters on the roads and greatly decreased the processing time required for clustering. Our tactic was to take only pairs of GPS points that likely came from a stationary GPS logger. Intuitively, we know that points measured from the same location are likely to be closer together than points measured from different locations. Our aim was to find a principled way to detect such pairs of temporally adjacent points measured from the same location. They key was to find a distance threshold: only pairs of points whose distance between is less than this threshold would be retained as likely coming from a non-moving logger.

To identify such points, we modeled GPS noise as a two-dimensional Gaussian, which has precedent in the GPS community [23]. We estimated the standard deviation of the error in our GPS loggers at approximately $\sigma = 4$ meters. Because of this error, multiple measurements from a non-moving logger change from one sample to the next. We can model the the distance between two points measured from the same location. The probability distribution of the distance between two samples taken from a normal distribution is called the normal difference distribution. For one-dimensional Gaussians, this distribution has a closed form [24]. Shaikh and Kitagawa [25] give a formula for the cumulative probability distribution of the distance between two, two-dimensional points $\boldsymbol{a}$ and $\boldsymbol{b}$ with different Gaussian distributions:

$$\boldsymbol{a} \sim N\left(\boldsymbol{\mu}_a, \begin{bmatrix} \sigma_{a,x} & 0 \\ 0 & \sigma_{a,y} \end{bmatrix}\right) \quad \boldsymbol{b} \sim N\left(\boldsymbol{\mu}_b, \begin{bmatrix} \sigma_{b,x} & 0 \\ 0 & \sigma_{b,y} \end{bmatrix}\right)$$

The cumulative probability distribution of the distance between these two random points is:

$$P(|\boldsymbol{a} - \boldsymbol{b}| < d) = c \int_0^d \int_0^{2\pi} \exp\left\{-\left(\frac{(r\cos\theta - \alpha_x)^2}{2(\sigma_{a,x}^2 + \sigma_{b,x}^2)} + \frac{(r\sin\theta - \alpha_y)^2}{2(\sigma_{a,y}^2 + \sigma_{b,y}^2)}\right)\right\} r\,dr\,d\theta$$

where

$$c = \frac{1}{2\pi\sqrt{(\sigma_{a,x}^2 + \sigma_{b,x}^2)(\sigma_{a,y}^2 + \sigma_{b,y}^2)}}$$

$$\alpha_x = \mu_{a,x} - \mu_{b,x}$$

$$\alpha_y = \mu_{a,y} - \mu_{b,y}$$

While we do not know of a closed form solution for this integral, we can fortunately simplify it. Since we are modeling a stationary GPS logger, we have $\boldsymbol{\mu}_a = \boldsymbol{\mu}_b$. And since the logger's noise characteristics are isotropic and do not change between sample points $\boldsymbol{a}$ and $\boldsymbol{b}$, we have $\sigma_{a,x} = \sigma_{a,y} = \sigma_{b,x} = \sigma_{b,y} = \sigma$. With these simplifications, we get

$$P(|\boldsymbol{a} - \boldsymbol{b}| < d) = 1 - \exp\left(-\frac{d^2}{4\sigma^2}\right)$$

This equation gives the probability that the distance between two sampled points from a stationary GPS logger will be less than d. As shown in Figure 4, when we set this probability to a high value, say 0.95, then $d = 13.85$ meters when $\sigma = 4$. Thus, with 95% probability, any pair of points measured from the same location will be within 13.85 meters of each other. Stated differently, taking all temporally adjacent pairs of points that are less than 13.85 meters apart will theoretically give a recall rate of 0.95 when we are looking for points from a non-moving logger. We use this in our clustering by filtering out temporally adjacent points whose distance apart is greater than this threshold. We note that a

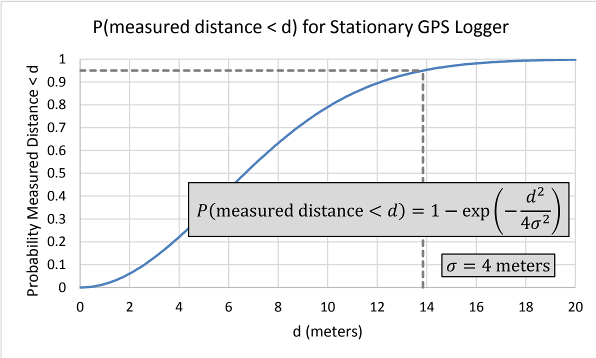

Figure 4: This is the cumulative probability distribution for the distance between two GPS measurements from the same location. There is a 95% chance the two points will be within 13.85 meters of each other.

vehicle moving at 10 miles per hour, sampled at our 10-second interval, would move about 45 meters between points, so even a slow-moving vehicle will produce inter-point distances greater than our threshold.

With these two changes (artificially adding departure points and filtering out moving points), we apply the regular agglomerative clustering described above. For purposes of our user study, we retained the top 30 clusters we found when sorted by time spent in each cluster.

Ground Truth Labels and Results

With each subject's clusters computed, we returned to them with an interactive map program for labeling their home and work clusters, show in Figure 5. Based on this labeling from our subjects, we verified that our clustering program found clusters representing all the subjects' homes and main work locations.

To test Placer, we used each subject's GPS data to compute a feature vector for each visit to each cluster. We set the visit arrival time at each cluster as the point when we first saw a GPS point within 200 meters of the cluster center, and the departure time as when the logger was subsequently 200

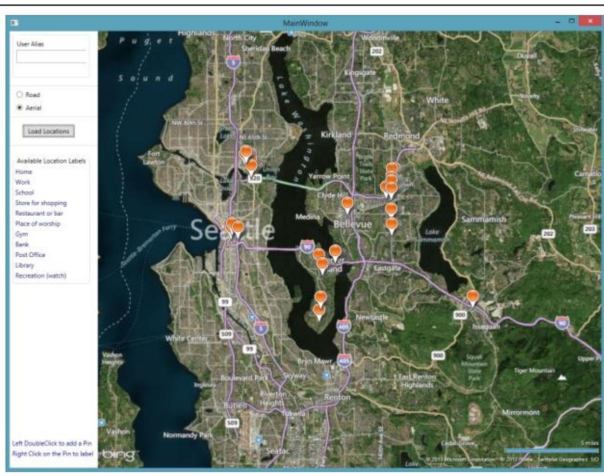

Figure 5: This is a screenshot of the program our subjects used to label their clusters with "home" and "work".

meters away. Each home and work cluster had multiple visits. The feature vector for each visit comprised the same 69 features that we used for our PSRC test above: temporal, demographic, and nearby businesses. We trained a decision tree on all the PSRC data using the same learning parameters as previously. The classifications resulted in potentially multiple inferred labels for each cluster, one for each visit. We declared the label correct if the majority of the inferred labels matched the actual label.

We were able to correctly label the home locations of all 28 subjects. We correctly labeled at least one work cluster of 18 of the 28 subjects (64%). The accuracy for labeling work is lower than we would have expected for the PSRC model. We attribute this to two factors. One is that a few of our subjects split their time between multiple workplaces on the campus of our institution, often within the same day, which split their visits unnaturally. The other factor is that our compensation for GPS warm-up is inadequate. Most of our participants park their vehicles in a parking garage while at work, leading to the warm-up problem. This is in contrast to their homes, where it is generally easier to get GPS data while parked. This is reflected in the fact that we found homes more successfully than workplaces. We were gratified, however, that we could train a classifier from relatively clean diary data and successfully test on GPS visit data taken under very different circumstances.

THE PROMISE OF SEQUENCES

It is natural to assume that people follow certain sequences of visits rather than picking their next location completely at random. For instance, we might expect most people go directly home after work. If this bias is true, then it can be exploited to help label a sequence of visits.

This idea was explored with data from four users with a conditional random field algorithm in [10] and from one user (apparently) using a hidden Markov model in [11]. With a large amount of data available from diary studies, we were curious to see if sequence analysis would be useful for a much larger sample of subjects.

We explored this by creating a simple first-order Markov model for place visits from our ATUS data. Specifically, we looked at 14 different ATUS place types. For each, we tracked the next places visited, giving an estimate of the Markov probability of next visiting a certain place category given the current place category. We ignored all of ATUS's transportation categories, because we found that transportation usually intervenes between actual place visits, making it uninteresting to consider. The Markov probabilities are show in Table 5.

An interesting aspect of this table is a destination of "home" dominates every other place category no matter where someone is currently visiting, other than already being home. For someone already at home, the two most popular places to go are "Store for Shopping" (21%) and "Other's Home" (18%).

To get an idea of the usefulness of sequence analysis, we can imagine that we somehow know a person's current place category. Our task is to label the next place they go. Using the maximum probabilities in each row of the Markov probabilities in Table 5, we would always label the next place as "Home", unless the current place was "Home", and then we would label the next place as "Store for Shopping". If we did this on the original ATUS data, we would correctly label the next place 43% of the time. This shows that there is worthwhile potential in considering the sequence of visits.

From Place \ To Place	Home	Work	Other's Home	Restaurant or Bar	Place of Worship	Store for Shopping	School	Outdoors	Library	Other Place	Bank	Gym	Post Office	Unspecified Place
Home	0.00	0.15	0.18	0.10	0.05	0.21	0.05	0.08	0.00	0.14	0.00	0.01	0.00	0.02
Work	0.68	0.00	0.03	0.10	0.00	0.05	0.03	0.02	0.00	0.06	0.00	0.00	0.00	0.01
Other's Home	0.56	0.05	0.00	0.08	0.03	0.11	0.02	0.04	0.00	0.08	0.00	0.00	0.00	0.01
Restaurant or Bar	0.37	0.12	0.10	0.00	0.05	0.14	0.03	0.04	0.00	0.14	0.01	0.00	0.00	0.01
Place of Worship	0.82	0.01	0.06	0.02	0.00	0.03	0.01	0.01	0.00	0.02	0.00	0.00	0.00	0.01
Store for Shopping	0.50	0.08	0.09	0.10	0.03	0.00	0.02	0.03	0.00	0.12	0.01	0.00	0.01	0.01
School	0.71	0.07	0.05	0.03	0.00	0.04	0.00	0.03	0.00	0.05	0.00	0.00	0.00	0.01
Outdoors	0.65	0.06	0.08	0.06	0.01	0.06	0.02	0.00	0.00	0.05	0.00	0.00	0.00	0.00
Library	0.47	0.04	0.06	0.05	0.01	0.13	0.07	0.03	0.00	0.09	0.02	0.01	0.01	0.01
Other Place	0.60	0.08	0.08	0.08	0.01	0.08	0.02	0.02	0.00	0.00	0.01	0.00	0.00	0.01
Bank	0.41	0.16	0.05	0.04	0.01	0.14	0.03	0.01	0.00	0.10	0.00	0.00	0.03	0.00
Gym	0.72	0.10	0.03	0.02	0.00	0.04	0.03	0.03	0.00	0.02	0.00	0.00	0.00	0.01
Post Office	0.44	0.10	0.04	0.04	0.01	0.16	0.03	0.02	0.01	0.08	0.06	0.01	0.00	0.01
Unspecified Place	0.62	0.07	0.07	0.06	0.02	0.07	0.02	0.02	0.00	0.06	0.00	0.00	0.00	0.00

Table 5: These are the computed Markov probabilities giving the probability of visiting a given place category after visiting another place category.

SUMMARY AND DISCUSSION

We have presented Placer, a technique for inferring semantic place labels, trained on data from two large diary surveys, ATUS and PSRC. We showed how both diary studies gave classification accuracies of around 0.73. With the PSRC data, we showed that omitting geographic features reduced the accuracy by about 14 percentage points. We showed how geographic features help label less frequently visited places such as "Eat Out", "Religious/Community" and "Shopping". We also tested our technique on GPS data from 28 subjects, successfully labeling all their home locations and 64% of their work locations. Our GPS clustering introduced a new technique for detecting places where the GPS logger was not moving.

The problem of labeling locations comes in multiple types. In our work with diary surveys, our goal was to correctly label each visit. This might be useful for a mobile application that needs to infer that a person is visiting a restaurant or movie theater, possibly for the first time, so it can take appropriate action. Our GPS study was slightly different in that we used the inferred labels on visits to label locations as either home or work. Once these familiar places are labeled, then there is no need for making sophisticated inferences, because geographic proximity to the labeled location is enough.

One limitation of using diary surveys to label places is that we are constrained by the survey's taxonomy of places. We may want places that were not included in the surveys, like camp ground, cemetery, or park. The surveys also necessarily aggregate different places that we may want to distinguish. For instance, PSRC's "Recreation – Watch" category could include movie theaters, sports stadiums, and opera houses.

The taxonomy of the survey may also be mismatched with our geographic business categories. For simplicity, we chose to use 15 high-level categories that covered all the businesses in our database. However, there are likely more targeted subsets of businesses that are especially useful for inferring certain place types. For instance, the place type "Restaurant or bar" is probably a strong function of the presence of a restaurant or bar, not just our "Food & Dining" category that also includes grocery stores.

There is also a problem of unifying taxonomies between multiple diary surveys. We did this in a simple way, but there may be more sophisticated approaches based on looking at the actual visit data.

The same place should sometimes be labeled differently for different people. For instance, a school could be someone's school or someone's workplace. Placer does this by examining an individual's demographics and visits to each place, allowing the same place to have different labels for different people. More subtlety, the same place may have different roles depending on the time of day. While we did not explore this, the PSRC data does make this distinction explicitly in one case by including labels "Home – Paid Work" and "Home – Other". This is an interesting direction for future work as it suggests we want to infer activities as well as place labels. The ATUS data supports this by including both a place label and an activity for each visit. Partridge and Golle [3] used the ATUS data to explore this connection, and it would be interesting to go further and label geographic places with activities.

Other promising future work should include an examination of the necessary intensity of location sampling. We sampled GPS at a 10-second interval. For battery efficiency, a longer interval may be sufficient, or an adaptive interval, or more efficient but less accurate measurements from WiFi.

Finally, it would be instructive to explore the advantages of analyzing sequences of place labels. As we showed, there is strong structure in the Markov probabilities.

REFERENCES

1. Miluzzo, E., et al., *CenceMe - Injecting Sensing Presence into Social Networking Applications* in *2nd European Conference on Smart Sensing and Context (EuroSSC 2007)*2007, Springer-Verlag. p. 1-28.

2. Ye, M., et al., *On the Semantic Annotation of Places in Location-Based Social Networks*, in *The 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2011)*2011. p. 520-528.
3. Partridge, K. and P. Golle, *On Using Existing Time-Use Study Data for Ubiquitous Computing Applications*, in *10th International Conference on Ubiquitous Computing (UbiComp 08)*2008, ACM. p. 144-153.
4. *American Time Use Survey*. Available from: http://www.bls.gov/tus/.
5. Zhu, Y., Y. Sun, and Y. Wang, *Nokia Mobile Data Challenge: Predicting Semantic Place and Next Place via Mobile Data*, in *Mobile Data Challenge 2012 Workshop*2012.
6. Kim, D.H., K. Han, and D. Estrin, *Employing User Feedback for Semantic Location Services* in *13th International Conference on Ubiquitous Computing (UbiComp 2011)*2011, ACM. p. 217-226.
7. Phithakkitnukoon, S., et al., *Activity-Aware Map: Identifying Human Daily Activity Pattern Using Mobile Phone Data*, in *First International Workshop on Human Behavior Understanding (HBU 2010)*2010, Springer-Verlag. p. 14–25.
8. Wolf, J., R. Guensler, and W. Bachman, *Elimination of the Travel Diary: An Experiment to Derive Trip Purpose From GPS Travel Data.* Transportation Research Record: Journal of the Transportation Research Board, 2001. 01(3255): p. 125-134.
9. Xie, R., et al., *Pattern Mining, Semantic Label Identification and Movement Prediction Using Mobile Phone Data*, in *8th International Conference on Advanced Data Mining and Applications (ADMA 2012)*2012, Springer-Verlag. p. 419-430.
10. Lin Liao, D.F., Henry Kautz, *Extracting Places and Activities from GPS Traces Using Hierarchical Conditional Random Fields.* The International Journal of Robotics Research, 2007. 26(1): p. 119-134.
11. Chen, Z., et al., *A Supervised Learning based Semantic Location Extraction Method using Mobile Phone Data*, in *2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE 2012)*2012. p. 548-551.
12. Laurila, J.K., et al., *The Mobile Data Challenge: Big Data for Mobile Computing Research*, in *Mobile Data Challenge 2012 Workshop*2012.
13. Montoliu, R., A. Martínez-Uso, and J. Martínez-Sotoca, *Semantic place prediction by combining smart binary classifiers*, in *Mobile Data Challenge 2012 Workshop*2012.
14. Zhu, Y., et al., *Feature Engineering for Place Category Classification*, in *Mobile Data Challenge 2012 Workshop*2012.
15. Huang, C.-M., J.J.-C. Ying, and V.S. Tseng, *Mining Users' Behaviors and Environments for Semantic Place Prediction*, in *Mobile Data Challenge 2012 Workshop*2012.
16. *Puget Sound 2006 Household Activity Survey*. Available from: http://www.psrc.org/data/surveys/2006-household/.
17. Friedman, J.H., *Greedy Function Approximation: A Gradient Boosting Machine.* Annals of Statistics, 2001. 29(5): p. 1189-1232.
18. Marmasse, N. and C. Schmandt, *Location-Aware Information Delivery with ComMotion*, in *Second International Symposium on Handheld and Ubiquitous Computing*2000. p. 157-171.
19. Ashbrook, D. and T. Starner, *Learning Significant Locations and Predicting User Movement with GPS*, in *Sixth International Symposium on Wearable Computers (ISWC 2002)*2002. p. 101 - 108.
20. Kang, J.H., et al., *Extracting Places from Traces of Locations*, in *2nd ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots (WMASH 2004)*2004. p. 110-118.
21. Hightower, J., et al., *Learning and Recognizing the Places We Go*, in *7th International Conference on Ubiquitous Computing (UbiComp 2005)*2005, Springer. p. 159-176.
22. Caoy, X., G. Congy, and C.S. Jensen, *Mining Significant Semantic Locations From GPS Data*, in *Proceedings of the VLDB Endowment* 2010. p. 1009-1020.
23. Diggelen, F.v., *GNSS Accuracy: Lies, Damn Lies, and Statistics*, in *GPS World*2007. p. 26-3.
24. Weisstein, E.W. *Normal Difference Distribution*. Available from: http://mathworld.wolfram.com/NormalDifferenceDistribution.html.
25. Shaikh, S. and H. Kitagawa, *Distance-Based Outlier Detection on Uncertain Data of Gaussian Distribution*, in *14th Asia-Pacific Web Conference (APWeb 2012)*2012, Springer. p. 109-121.