An Experiment in Discovering Personally Meaningful Places from Location Data

Changqing Zhou, Pamela Ludford, Dan Frankowski, Loren Terveen
Department of Computer Science and Engineering
University of Minnesota
200 Union ST SE, 4-192, Minneapolis, MN 55414
{czhou, ludford, dfrankow, terveen}@cs.umn.edu

ABSTRACT
As mobile devices become location-aware, they offer the promise of powerful new applications. While computers work with physical locations like latitude and longitude, people think and speak in terms of places, like "*my office*" or "*Sue's house*". Therefore, location-aware applications must incorporate the notion of places to achieve their full potential. This requires systems to acquire the places that are meaningful for each user. Previous work has explored algorithms to discover personal places from location data. However, we know of no empirical, quantitative evaluations of these algorithms, so the question of how well they work currently is unanswered. We report here on an experiment that begins to provide an answer; we show that a place discovery algorithm can do a good job of discovering places that are meaningful to users. The results have important implications for system design and open up interesting avenues for future research.

Author Keywords
Ubiquitous computing, location-aware applications, clustering algorithms, place discovery, field studies.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
Location-aware applications use technology like the Global Positioning System (GPS) to compute the physical location of a mobile device. There has been much research in this area [7, 8] and an increasing number of commercial systems(e.g., uLocate.com). While these systems operate mostly in terms of physical locations (e.g., latitude/longitude), we argue that realizing the full potential of this class of applications requires them to incorporate a more abstract notion of place. Research in environmental psychology has shown that people naturally structure their experience around socially meaningful places - home, office, school, church, coffee shop, pub, etc [2]. Places derive their meaning from social conventions concerning expected and unexpected activities, their private or public nature, possibilities for communication, etc [3, 5].

How to acquire a user's personally meaningful places thus becomes an important research topic in place-based systems. Wilenmann and Leuchovius argue that location-based services should describe location in ways relevant to users, such as "I'm home" [12]. ComMotion [8] pioneered an interactive discovery approach to acquiring user locations. The comMotion device constantly takes GPS readings and takes the loss of signal as a significant cue that the user has entered a building. Ashbrook and Starner [1] applied a clustering algorithm to group GPS readings into"significant locations". Other researchers have explored the use of methods such as Markov models and Bayesian networks to learn significant locations and probable transitions between them [6, 10].

Prior work has provided compelling proof-of-concept implementations of place discovery algorithms. However, we know of no quantitative empirical studies that evaluate how well any of these algorithms actually do their job, i.e., how accurately they discover a user's personal places. We conducted such a study, establishing a performance baseline so that designers can know what to expect from an algorithm and how to incorporate it in an interactive place discovery system. The key research questions we studied were:

> *Can an algorithm accurately discover users' places from records of their physical locations? Further, can the algorithm discover the places that most matter to users?*

In our experiment, we equipped subjects with GPS-enabled mobile phones to collect location data over a three week period; subjects also kept a daily log of the places they visited. We then applied a place discovery algorithm [13] to the location data, generating a map for each subject that showed their discovered places as points on the map. Finally, we carried out a semi-structured interview organized around each subject's logged places and the generated map to evaluate the accuracy of the discovery algorithm.

CHI 2005, April 2-7, 2004, Portland, Oregon, USA.
ACM 1-59593-002-7/05/0004.

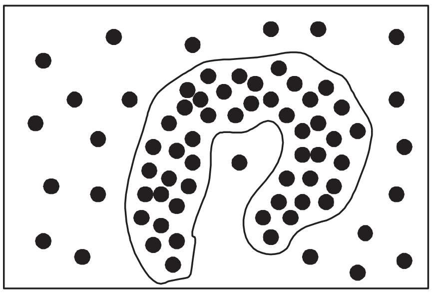


Figure 1. Clusters are formed where density is high.

BACKGROUND

The DJ-Cluster discovery algorithm [13] takes a database of location readings - a personal location dataset - as input and produces a list of places as output. Each reading represents a time-stamped location, for example:

> *latitude 44°57.02', longitude 93°15.48', 6:05pm Friday, September 10, 2004.*

DJ-Cluster is a density-based clustering algorithm; that is, it forms clusters of raw location readings where the point density is high (see Fig. 1). These clusters are the candidate places for a user.

While the overall discovery process is interactive, the algorithm still should be as accurate as possible: most of the places users care about should be found, and users shouldn't have to deal with a bunch of spurious candidates. We therefore need a systematic way of measuring an algorithm's performance. We use the following method and metrics (taken from [13]).

Evaluating a Place Discovery Algorithm:

1. Collect subjects' personal location dataset.
2. To establish a baseline for evaluating system accuracy, have subjects log the places they visit during the data collection period (call this set *B* for *Baseline*).
3. Run the algorithm on the personal location dataset to discover candidate places for each subject (call this set *D* for *Discovered*).
4. Present the discovered places to each subject on a map. Ask subjects to match their baseline places to the discovered ones. Of the discovered places that weren't in the baseline, ask the subject whether they represent personally meaningful places (as opposed to a location they can't recognize at all, or a meaningless location such as an area of road where they were stuck in traffic). Call this set of places *DM*, for *Discovered Meaningful*.

   The matching process results in the following sets, summarized in Fig.2:

   - *BD*: Baseline places that were discovered = *B* ∩ *D*.
   - *BN*: Baseline places not discovered = *B* - *D*.
   - *DM*: Discovered places that are meaningful.

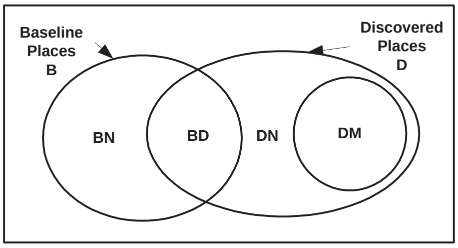
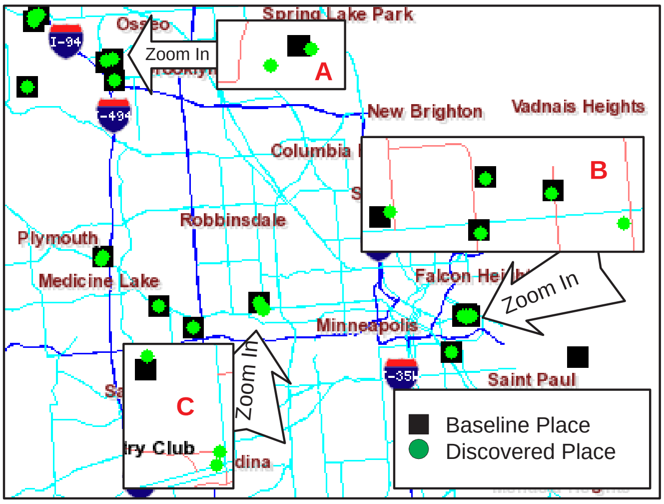


Figure 2. Sets of baseline and discovered places.



A: Community center: swimming pool and ice arena
B: Campus: parking I, parking II, EE/CS, coffee shop, and Chipotle
C: Work: work and 2 traffic stops

Figure 3. Baseline and discovered places for one person.

5. We quantify the performance of the algorithm using three metrics: *Precision*, *Recall*, and *SurpriseFactor*:

$$Precision = |BD|/|D|$$

$$Recall = |BD|/|B|$$

$$SurpriseFactor = |DM|/|D|$$

We offer several observations. First, the sum of the *Precision* and *SurpriseFactor* scores is an interesting derived quantity. It tells how many of the discovered places were"good," i.e., either in the subject's baseline or discovered and judged meaningful. Second, not all places that are meaningful to a person are equally important: places like one's *home*, place of *work*, or child's *school* may be more important than a *friend's house* or *restaurant* one visits occasionally. Therefore, it is useful to obtain users' judgments of the important places, and then compute versions of these metrics for important places only. Fig. 3 illustrates these concepts with real user data. The squares represent baseline places (*B*), and the circles represent places discovered by the algorithm (*D*).

THE EXPERIMENT

Subjects. Our intuition was that people's daily activities, and thus the types of places they visited, would depend on their life stage. For example, we expected 20 year old college students, 40 year old working parents, or 60 year old retirees to frequent different types of places. Therefore, we recruited subjects from across this spectrum.

We ended up with 28 subjects, all from a major U.S. metropolitan area. Some live in a core city, some in the suburbs. They used a variety of travel modes, including walking, biking, public transportation, and personal car. Their ages ranged from the early 20s to late 60s, with an average in the early 30s. Twenty were male, 8 female. Three subjects had preschool children, 4 had school aged children, and 2 had adult children not living with them. They were highly educated, with 2/3 having college or advanced degrees. They included 6 college students, 4 engineers, 4 information technology professionals, 4 teachers, a range of other professional and service jobs, as well as several retired people. Since this study is formative research, aimed at observing subject experiences and using the results to guide more traditional empirical work in the future, we believe our subject pool is large enough for the study.

Data collection. We equipped subjects with a GPS-enabled Motorola i88s cell phone with Nextel service that ran the Accutracking software (www. accutracking.com). We set the software to take a GPS reading every minute and send it to a server for storage. Subjects carried the phone for about three weeks, and were instructed to keep the phone with them and turned on at all times. However, they could turn off the phone for privacy reasons whenever they wanted. Subjects also kept a diary [4, 11] of the places they visited each day. They received a daily reminder (email, instant message or phone call); they could return their list of places via email or record them in a notebook. We met with each subject at the beginning of the study to give them the phone, demonstrate its use, and instruct them in the data collection procedure. And after data collection was complete, we conducted a semi-structured interview with each subject.

Interview preparation. We ran our discovery algorithm on each subject's personal location dataset. For each subject, we printed out an overview map showing all the discovered places, and a set of more detailed maps showing nearby groups of discovered places at a higher level of resolution; for example, at the detailed level, street names were visible, and the city block of each place was apparent. We also printed a table of each subject's baseline places (Table 1).

Interviews. We conducted a semi-structured interview with each subject organized around the maps and table. We first led subjects through the process of matching their baseline places (in the table) to the discovered places (on the map). We had assigned a label (A-Z, then A1, A2, ...) to each discovered place; when a subject found a discovered place that corresponded to a baseline place, we wrote the label in the *Label* column in Table 1. This resulted in the various sets defined in the evaluation procedure and let us compute the *Precision*, *Recall*, and *SurpriseFactor* metrics. Further, since some places matter more than others, we asked subjects to rate the importance of each place on a scale of 1 (least important) to 5 (most important). We recorded this in the *Importance* column of the table.

Description	Label	Importance
Home	A	5
Office	F	5
Cooke Hall	F	5
Ted's House	I	4
Radisson MNPLS		1
Brits Pub		1
Newsroom		1
Blockbuster		1
Mom and Dad's		5

Table 1. Excerpt from the interview table

Statistics	Readings	Data collect days	Mean readings per day
Total	152,741	516	7,008
Mean per subj.	6,364	22	292
Std Dev	3,997	3	167

Table 2. Personal Location Dataset

THE DATASET

All 28 subjects logged their data for three weeks. Time constraints kept us from scheduling interviews for 3 subjects, so we discarded their data. Also, the interview with one subject was unsuccessful, so we discarded this data as well. Thus, we ended up with data for 24 subjects.

Personal Location Dataset. Table 2 summarizes the location readings collected from the subjects. This is a large amount of real data - we know of no other studies with samples of this size. Subjects followed the data collection procedure quite faithfully. The average number of readings per subject - 6,364 - represents over 100 hours of data.

Baseline and Discovered Places. As shown in Table 3, subjects logged a total of 681 places in their diaries during the experiment, or about 28.4 per subject. The system discovered 56 non-baseline places; for each subject, it discovered at least some meaningful places that the subject had forgotten.

RESULTS

Table 4 presents the results of our three main evaluation metrics. For each, we computed the metric for all places as well as just those places that were important to subjects (we

Statistics	Baseline	Out of town	Non-baseline discovered
Total	681	29	56
Mean per subj.	28.4	3.2	2.3
Std Dev	13.0	2.8	2.8

Table 3. Statistics about places for all subjects.

Statistics	Recall	Precision	Surprise Factor
All Places	48%	85%	14%
Important Places	76%	35%	2%

Table 4. Quantitative Results

counted a place as important if subjects scored it a 4 or 5). Let's briefly consider the meaning of these results.

First, what one would most like to know is: are these numbers good or bad? Unfortunately, since this is the first empirical evaluation of a place discovery system that we know of, there is no point of comparison. Therefore, we see these numbers as establishing a baseline that other systems can use as a point of comparison. By providing an evaluation framework and metrics and publishing these numbers, we give future researchers in the area a way to answer the question: is this "better" or just "different"? [9]

Second, observe that the algorithm discovers personal places with high precision and low recall. Since *Precision* and *SurpriseFactor* sum to 99%, nearly all the places discovered by the algorithm were personally meaningful to subjects. On the other hand, the algorithm failed to discover just over 50% of subjects' baseline places. If we recall that the system is intended to be an interactive aid to a user, then we can say that what it tells a user will almost always be useful, but that it will make less than half of all useful suggestions.

Why is recall so low? Through asking users about their baseline places that weren't discovered, it is our belief that most of these actually were only visited once or twice during the data collection period. For example, one subject's undiscovered places included *City Library, Home Depot, Lifetime Fitness Center, two lunch restaurants, Science Museum and a friend's house*. He reported visiting each of these places just once during the experiment. This would mean that the readings generated for that place would be too few too satisfy the density requirements for the discovery algorithm.

Third, the performance of the algorithm looks quite different when we consider only places that subjects judged important. The system discovered 138 of 181 important baseline places, or 76%. Precision of important places seems low: only 37% of the discovered places were important. However, the percentage of baseline places that were important was even lower - just 25%. Therefore, we can offer an interpretation that unifies the numbers for all places and important places:

> *The places the algorithm discovers are quite likely to be meaningful to a user, and places that are important to a user are likely to be discovered.*

CONCLUSIONS

We performed an experimental evaluation of a place discovery system. We found that an algorithm can do a good job of discovering places that are meaningful to users. The places the algorithm discovers are quite likely to be meaningful to a user, and places that are important to a user are likely to be discovered. We also found that certain places that were meaningful to users were not discovered; we speculate that this was because users did not visit them often during the data collection period. These results have important implications for system design and raise interesting research challenges. For example, we are planning to investigate whether our data will support automatic prediction of place importance, e.g., based on how frequently a place is visited, how much time is spent there, or whether it is visited at regular times.

REFERENCES

1. D. Ashbrook and T. Starner. Learning significant locations and predicting user movement with GPS. In *Proc. IEEE Sixth International Symposium on Wearable Computing*, 2002.
2. K. A. Franck and L. H. Schneekloth, editors. *Ordering space: types in architecture and design*. Van Nostrand Reinhold, 1994.
3. R. Genereux, L. Ward, and J. Russell. The behavioral component in the meaning of places. *Journal of Environmental Psychology*, 3:43–55, 1983.
4. J. Hightower. From position to place. In *Proc. Workshop on Location-Aware Computing*, 2003.
5. B. Kramer. Classification of generic places: Explorations with implications for evaluation. *Journal of Environmental Psychology*, 15:3–22, 1995.
6. L. Liao, D. Fox, and H. Kautz. Learning and inferring transportation routines. In *Proc. AAAI*, 2004.
7. S. Long, R. Kooper, G. D. Abowd, and C. G. Atkeson. Rapid prototyping of mobile context-aware applications: The cyberguide case study. In *Proc. MobiCom*, pages 97–107, 1996.
8. N. Marmasse and C. Schmandt. Location-aware information delivery with commotion. In *Proc. HUC*, pages 157–171, 2000.
9. M. Naaman, Y. J. Song, A. Paepcke, and H. Garcia-Molina. Automatic organization for digital photographs with geographic coordinates. In *Proc. JCDL*, 2004.
10. D. Patterson, L. Liao, D. Fox, and H. Kautz. Inferring high-level behavior from low-level sensors. In *Proc. UbiComp*, 2003.
11. J. Rieman. The diary study: a workplace-oriented research tool to guide laboratory efforts. In *Proc. CHI*, 1993.
12. A. H. Weilenmann and P. Leuchovius. I'm waiting where we met last time: exploring everyday positioning practices to inform design. In *Proc. NordiCHI*, 2004.
13. C. Zhou, D. Frankowski, P. Ludford, S. Shekhar, and L. Terveen. Discovering personal gazetteers: An interactive clustering approach. In *Proc. ACMGIS*, 2004.