Predicting Location-Sharing Privacy Preferences in Social Network Applications

Greg Bigwood[1], Fehmi Ben Abdesslem[2], and Tristan Henderson[1]

[1] School of Computer Science, University of St Andrews, UK
{gjb4,tnhh}@st-andrews.ac.uk
[2] Computer Laboratory, University of Cambridge, UK
fb375@cam.ac.uk

Abstract. The prevalence of social network sites and smartphones has led to many people sharing their locations with others. Privacy concerns are seldom addressed by these services; the default privacy settings may be either too restrictive or too lax, resulting in under-exposure or over-exposure of location information. One mechanism for alleviating over-sharing is through personalised privacy settings that automatically change according to users' predicted preferences. In this paper, we use data collected from a location-sharing user study ($N = 80$) to investigate whether users' willingness to share their locations can be predicted. We find that while default settings match actual users' preferences only 68% of the time, machine-learning classifiers can predict up to 85% of users' preferences. Using these predictions instead of default settings would reduce the over-exposed location information by 40%.

1 Introduction

Many popular social network sites (SNSs), e.g., Facebook Places[3], and Foursquare[4], enable people to share their real-world locations in real-time, for instance by using their smartphone's 3G data connection to upload details of their location. These location data may be sensitive, and as such should not be made public by default. Typical SNS privacy settings are such that when a user publishes their location information, the SNS uses a series of rules to determine who is given permission to view this location. The decision depends on the access control settings that the SNS provides for the data, as well as the default settings that the user has provided for the location. For example, Facebook Places allows users to restrict access to their location information using several mechanisms:

1. When a user posts an update to Facebook they can decide before they post it whether to include their location information.
2. When a fellow user indicates that they were co-located with another user, each of the users tagged as being co-located can independently decide whether to allow Facebook to publish their location information.

[3] http://www.facebook.com/places/
[4] http://www.foursquare.com/

3. A user can optionally be automatically tagged in location posts by their contacts. Users can also remove the ability of other users to indicate they were co-located.
4. In addition, users can select which of their contacts can see their location by selecting which *friend list* (a user-defined subset of contacts) has visibility of the location information, or by reverting to a user-defined default list.

The complexity of these rules, and the number of ways in which location data may be shared, means that users may be unable to specify, or unaware of, the appropriate group of individuals with whom they wish to share their location. They may even be unaware that their location is being shared at all. It may be the case, then, that users share their location with either too few people (under-exposure) or too many people (over-exposure). The latter is clearly an undesirable privacy leak, as highlighted by websites such as PleaseRobMe[5], which uses Twitter[6] and Foursquare location updates to determine the location of unoccupied homes.

Possible solutions to this problem include educating users about risks, or designing user interfaces and policy languages that allow users to express their actual preferences. This assumes, however, that users are firstly able to determine their correct preferences, and that these preferences are static. If users' preferences are highly variable, use of the application may become burdensome, as they will have to select their preferences each time they wish to publish some information on an SNS.

In this paper we investigate the following questions:

1. Can users determine their appropriate default location-sharing privacy settings?
2. Can these default settings be improved by predicting users' preferences?

Using data from a real-world user study, we demonstrate that users' day-to-day preferences for sharing their location do not always match their static default settings. We then test a range of different classifiers to explore whether we can predict their preferences.

2 Related work

The privacy risks of using location-based services (LBSs) and applications have been studied for some time, e.g., [1,6]. More recently, a Microsoft survey revealed that 84% of users are concerned about potential losses of privacy when using LBSs, and 49% would be more comfortable if they could easily and clearly manage who sees their location information.[7] The rise of SNSs that allow location updates has led to new privacy risks, as surveyed by Fusco et al. [9]. In particular, SNSs allow a users' locations to be leaked by their friends [16,23]. It has also been shown that users often inadvertently reveal their locations on SNSs such as Twitter [13].

Solutions to alleviating location privacy risks include trusted servers that deliver location updates in a manner that offers k-anonymity [11], cloaking or "confusing"

[5] http://www.PleaseRobMe.com/

[6] http://www.twitter.com/

[7] Microsoft Research Data Privacy Day 2011,
http://www.microsoft.com/presspass/features/2011/jan11/01-26dataprivacyday.mspx

a path by obfuscating locations [14,19], or new applications that provide users with precise privacy controls [24].

Our interest is in predicting location-sharing privacy preferences so as to reduce the burden on the user. Li et al. [18] and Toch et al. [26] attempt to create appropriate default privacy settings for users. As we show in this and previous work [2], accurate default settings are insufficient as users' preferences may vary with context. Corapi et al. [8] show how to build and update rules by observing user behaviour. Voulodimos et al. [27] use clustering methods to group users by location and preferences, which could be useful for predicting privacy preferences.

As well as SNS users sharing their own information with each other, aggregate data about social networks may also be shared, e.g., with researchers or advertisers. Bhagat et al. [7] use link prediction to preserve privacy in such shared data. Conversely, Narayanan et al. [20] use link prediction to deanonymise such data.

3 Methodology

We aim to address two questions relating to user location privacy in this work:

Can users set their default settings appropriately? In previous work [2] we have found that users have complex and sometimes conflicting rationales behind their decisions to share their location. This can make it difficult to create appropriate default privacy settings for users.

We use data from a location-sharing user study to determine whether users set their SNS privacy-sharing settings appropriately. By comparing the default sharing settings chosen by users with the settings that they choose *in situ* when actually sharing a location, we can compare users' expectations of their location-sharing behaviour with their behaviour in a real-world setting. This helps estimate how many privacy leaks may occur due to incorrectly specified user settings.

Can we predict user sharing behaviour? Second, we attempt to classify user behaviour, so that we can predict whether a user will share a particular location with particular groups of their SNS contacts. By accurately predicting a user's behaviour, it may be possible to reduce privacy leaks. A predictive application could alert a user when their behaviour deviates from a prediction, or could automatically restrict location-sharing to within the predicted parameters.

In order to predict location-sharing behaviour, we must decide what information is useful for predicting user location sharing. We use the same user-study data to determine what information is useful for predicting location sharing behaviour. We then go on to evaluate various prediction approaches.

3.1 Location-sharing study data

We previously conducted a location-sharing user-study, details of which can be found in [4]. 80 participants in two UK towns, London and St Andrews, carried a smartphone running a custom location-sharing application for one week.

The study comprised three stages of data collection:

Pre-briefing. Participants were asked to install a Facebook application on their personal Facebook accounts and to fill out a questionnaire. The questionnaire asked questions about Facebook usage (e.g., “Have you ever disclosed your location or activity on your Facebook status?”) and also asked participants for their default location-sharing settings for each of their Facebook *friend lists*.

Location-service usage. Over the course of the week, participants used the smartphones to share their locations on Facebook. Additionally, using the Experience Sampling Method [17], we periodically sent questions to the smartphones to ask participants about their context and whether, and with whom, they would be willing to share their location.

Debriefing. Using a semi-structured interview, participants provided more detailed explanations about their sharing choices.

We use these data to compare the default settings provided with the responses given during the location-service usage stage. We also use these data to drive a location-sharing predictor.

3.2 Dataset attributes

The complete dataset of the responses provided by participants, is available online [3]. To develop a classifier, we used the following attributes:

1. A unique identifier for each user.
2. The question asked. Four slightly different questions were used in the study, although the goal of all of the questions was to elicit location-sharing preferences for the context of the particular time of the question.
3. The time that the question was asked and answered $\{0700 - 1200, 1200 - 1400, 1400 - 1700, 1700 - 2100, 2100 - 0700\}$. This may be useful to see if the time of day that the user is present at a location affects whether the user shares their location.
4. The type of location {academic, food/drink, leisure, library, residential, retail} as inferred from GPS data and queries of various location databases including Yell and Google Maps.
5. A Facebook *friend list* ID, representing the group to which a user shared (or did not share) their location, and the size of the list.
6. A boolean indicating whether a user did share their location.
7. Co-location {no-one, friends, strangers}: users were asked whom they were with at the time of a question.
8. Four attributes from the pre-briefing data: whether a user had used their phone to share their location prior to the study; whether they had shared their location on Facebook; whether they had changed their Facebook privacy settings from the defaults; and whether they were a frequent mobile phone Facebook user.

Any locations where we were unable to determine a type were removed, leaving $4,033$ instances.

4 Evaluation

To correctly predict whether a user is willing to share their location, we need to determine whether the user would agree to share their current location with groups of their Facebook friends. To assess the accuracy of our prediction schemes, we compare the result of the prediction with the actual response that the user gave us as to whether they would share their location; i.e., if we predict that a user shared their location, and they did indeed share their location, this is considered a correct prediction. This provides a measure of the percentage of correctly-classified instances, and we use this *correct predictions* metric to compare a variety of machine-learning techniques.

To determine whether users can define appropriate default settings, we use the default privacy settings for each friend list provided in the pre-briefing, and compare these to responses given during the ESM study. Any differences where sharing took place with a group of users that was not actually desired, are considered *privacy leaks*. For instance, if a user had chosen default settings that shared photos with their family and friends, but in the course of the study shared photos only with their family, then this is a privacy leak. In terms of evaluating our prediction schemes, we desire a high *correct prediction* rate, and a low *privacy leak* rate.

As a baseline performance for our analysis, we use the default settings provided by users for each *friend list*. This shows how well users can predict their own sharing behaviour, and ideally our prediction schemes should do better than this.

4.1 Simple classifiers

We evaluate two methods for classifying location sharing. First, we use simple online classifiers which work on the data in chronological order to predict location-sharing preferences for the current point in time:

List ID (Li) Use the last response provided by the user for this Facebook *friend list*. When no previous response exists, use the default setting for this list.

Location (Loc) Use the last response provided for this location, regardless of *friend list*. When no previous response exists, use the default setting for this list.

List ID in Location (LiLoc) Use the last response provided for this list at this location. When no previous response exists, use the default setting for this list.

List ID < List ID in Location (Li < LiLoc) Use the last response provided for this list at this location. When no previous response exists, use the last response provided for this list. Otherwise, use the default setting for this list.

Location < List ID in Location (Loc < LiLoc) Use the last response provided for this list at this location. When no previous response exists, use the last response provided for this location. Otherwise, use the default setting for this list.

4.2 Machine-learning classifiers

Second, we use the data to drive machine-learning classifiers. We apply implementations of various classification algorithms from the Weka toolkit [28] to the dataset, and assess the performance of the classification approaches using ten-fold cross validation.

We use three algorithms with increasing computational complexity as the trade-off between performance and computation time is an important consideration for delivering timely predictions to users.

Naïve Bayesian (NB) The Naïve Bayesian classifier [15] assumes that factors of classification are independent from one another. NB is particularly suited to situations with lots of attributes, such as our input data.

J48 Classifier (J48) The J48 classification algorithm is an implementation of the C4.5 algorithm [21]. This classification algorithm builds a decision tree from the input data, and selects the classifications accordingly. J48 requires more time to compute than NB.

Rotation Forest (RF) The Rotation Forest scheme [22] analyses subsets of features to train the J48 algorithm, resulting in improved performance at the expense of increased computational complexity.

4.3 What input information is useful?

Accurate predictions of location sharing rely on having sufficient information to make an informed decision. If we can identify the information that is useful for predicting location sharing from the dimensions of our user-study data, we can reduce the amount of information needed in future prediction systems. This will be useful to researchers conducting future user studies, as it may guide the selection of questions to ask participants and decisions about what information should be collected. We therefore assess which combinations of input data give best classification performance.

We select the best features from the input data using the Attribute Selection Classifier (ASC), which uses the CfSubsetEval method [12]. One of the three machine-learning classifiers described in Section 4.2 is then applied to the selected features. We refer to these as ASC:NB, ASC:J48 and ASC:RF.

4.4 Can we minimise privacy leaks?

One danger of using an automated prediction technique when dealing with private data is that the predictor may cause a privacy leak, and share data the user did not wish to share. Our final optimisation is to attempt to preserve user privacy while at the same time maintaining accurate predictions, by minimising these privacy leaks.

We therefore use a Cost Sensitive classifier (CS), where privacy leaks are weighted as ten times worse than not sharing a location at all. Using this approach, the machine-learning algorithms try to avoid classifying instances that would result in privacy leaks. We use the CS classifier with the classifiers from Section 4.2 and refer to them as CS:NB, CS:J48 and CS:RF.

5 Results

We now apply the classification and learning techniques described in the previous section to our dataset, in order to investigate whether privacy settings are predictable.

5.1 Simple Classifiers

First, we present the results of the simple techniques. Table 1 shows that the simple classifiers perform better than using the default settings, which sets the correct privacy settings 68% of the time. If users rely solely on their default settings, however, they will over-share 10.71% of the time. Hence, although augmenting the default settings with the simple classifiers increases the correct predictions, it also increases the privacy leaks; for instance, using the last response for the chosen *friend list* (Li) increases privacy leaks from 10.71% to 12.77%. We could therefore use the simple approaches when we want increased accuracy, but use the default when we want fewer privacy leaks.

5.2 Machine-learning approaches

The results in Table 2 show that we can increase the correct prediction rate using machine-learning techniques. Both the RF and J48 classifiers perform better than using the default settings, and while NB has also a higher correct prediction rate than the default settings, it provides a greater rate of privacy leaks.

Note that the RF classifier performs particularly well, far outperforming the default settings both in terms of correct prediction and privacy leaks. Thus there is benefit in using machine learning in this way to classify location-sharing instances.

Classifier	Correct predictions (%)	Privacy leaks (%)
Default	68.00	10.71
Li	73.49	12.77
Loc	76.44	11.28
LiLoc	71.71	12.60
Li < LiLoc	73.59	13.12
Lo < LiLoc	75.77	12.55

Table 1: Prediction performance for simple classifiers.

Classifier	Correct predictions (%)	Privacy leaks (%)
Default	68.00	10.71
NB	75.18	14.88
RF	83.29	3.33
J48	76.72	10.46
ASC:NB	73.02	15.35
ASC:J48	75.60	9.94
ASC:RF	75.65	11.95

Table 2: Prediction performance for machine-learning classifiers.

5.3 Which are the most useful attributes for classification?

Now that we have seen that we can outperform the default settings, we consider how much information is needed to drive the machine-learning classifiers.

Weka's attribute selection classifiers indicate that two attributes are most important: the Facebook*friend list* ID and size. Table 2 shows that when using only these two items of information, we achieve classification performance better than the user defaults, but slightly poorer than using all of the attributes. Interestingly, other pieces of context such as location and time-of-day do not have an great impact on a user's decision to share. Instead, it appears that users prefer to share consistently with the same set of friends. This is a surprising result given the different behaviour found in other user studies [25].

5.4 Cost-sensitive classifiers

Table 3 presents results for the cost-sensitive classifiers. As intended, these produce far fewer privacy leaks than using the user defaults, with the cost-sensitive RF classifier producing a similar overall rate of correctly-classified instances to user defaults (67.49%) but with only 0.72% privacy leaks versus 10.71% for the defaults.

Classifier	Correct predictions (%)	Privacy leaks (%)
Default	68.00	10.71
CS:NB	67.07	2.16
CS:J48	50.48	0.00
CS:RF	67.49	0.72

Table 3: Prediction performance for cost-sensitive classifiers.

5.5 In practice

The previous results show that the privacy settings chosen by users are consistent enough to be predicted. The ten-fold cross validation used in the results previously presented is not chronological, however; it uses training sets containing privacy settings that are not necessarily anterior to the predicted privacy settings.

For our system to be used in practice, we envisage a prediction application that runs on a mobile device and would recommend privacy settings to users. Ideally, it would first use the users' default privacy settings for an initial learning phase, while observing how the users occasionaly override their privacy settings. Then, it would keep using the users' default settings and let them override these default settings, but could also alert them when it predicts privacy settings that are different from their settings.

To implement such an application on a mobile device, the prediction scheme must only rely on previous privacy settings, collected in the learning phase. Table 4 and 5 show the performance of classifiers when using only previously-collected privacy settings in the training sets.

We observe that depending on the classifier and the size of the training set, we still achieve better prediction performance than the default settings. The RF classifier again outperforms the NB classifier, achieving up to 86.49% of correct predictions, while keeping the privacy leaks rate at 6.47%. This is a reduction of almost 40% compared to the 10.71% of privacy leaks occuring when using the default settings.

Note that the size of the training set has an impact on the performance of the prediction. The larger the training set, the better the prediction should be, as it learns more about the users' privacy preferences. But there is also a risk of overfitting the model by adjusting to very specific random features of the training data. This explains the performance variations in our results, depending on the training set size.

Training set length (days)	Correct predictions (%)	Privacy leaks (%)
Default	68.00	10.71
1	67.64	20.58
2	72.46	16.85
3	70.48	18.16
4	67.65	21.48
5	66.69	21.13
6	67.35	19.62

Table 4: Performance of chronological prediction using the NB classifier.

Training set length (days)	Correct Predictions (%)	Privacy leaks (%)
Default	68.00	10.71
1	84.63	7.72
2	86.49	6.47
3	85.62	6.76
4	84.61	7.18
5	84.04	7.17
6	83.78	7.10

Table 5: Performance of chronological prediction using the RF classifier.

5.6 Discussion

Figure 1 summarises the performance of the various classifiers. The outlined region to the upper left of the default baselines indicate where classifiers that perform better (both in terms of classification accuracy and preventing privacy leaks) should lie. It can be seen that the simple classifiers can outperform the default settings in terms of correct predictions, even though privacy leaks are higher. If the accuracy of the predictions is of most importance, then the RF classifier offers the best performance. Conversely, if minimising privacy leaks is more important, then using a cost-sensitive RF classifier gives similar performance to using user-selected defaults, but with considerably fewer privacy leaks.

The RF, J48, and ASC:J48 machine-learning classifiers all provide better performance in terms of both prediction accuracy and privacy leaks. The performance of ASC:J48 also shows that it is possible to make location-sharing predictions with better performance to the default settings using only the Facebook *friend list* size and ID. This has interesting implications for the design of privacy user interfaces for SNSs: rather than asking for a single set of privacy defaults as at present, it might be sufficient to ask users to set privacy preferences for each of their groups of friends, or at least to use this to bootstrap a learning system.

6 Conclusion

This paper has presented an initial investigation into the predictability of users' location-sharing privacy preferences in mobile social networks. Using data from a user study of 80 participants, we find that a user's predetermined default setting reflects their *in situ* preferences only 68% of the time. It is possible to improve upon this using machine-learning techniques, both improving accuracy and decreasing the rate of privacy leaks. In particular, a rotation forest classifier can provide 86% accuracy.

The attribute selection classifier indicates that a machine-learning classifier can also achieve similar performance to the user defaults using only two pieces of information for prediction. Surprisingly this information relates to the group with which the information is being shared, and other contextual information such as location and time of day do not appear to be important.

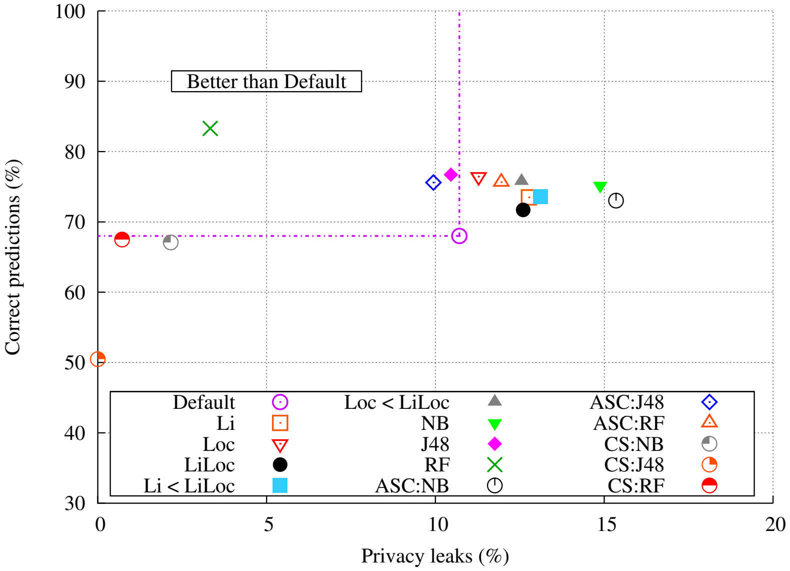

Fig. 1: A scatterplot showing the relative performance of the classification approaches. *Li* = Facebook *friend list*. *Loc* = Location. *LiLoc* = Facebook *friend list* in Location.

This work is still in its preliminary stages. Plans for future work include further analysis as to which subsets of user data are most useful for predicting location sharing. In particular, perhaps time of day is more important than indicated here? We also need further investigation into the relative weight of a privacy leak in comparison to being overly restrictive of location sharing. Perhaps the cost of a privacy leak should be considered to be dynamic, e.g., depending on some measure of the sensitivity of the specific locations being shared.

We currently assume that the predictor will run on a user's device, and so need to compare the complexity and power-efficiency of the various classifiers if we are to conserve energy on a mobile device. But in doing so, we have also only looked at each user in isolation. Distributed predictors might be more effective, e.g, using colocated users' privacy preferences to help predict privacy preferences for a given location. In effect, this could be considered "crowdsourcing" a predictor: if people near me do not wish to share their presence in this current place, then I may be likely to consider this place to be private.

More broadly, in addition to helping us test new hypotheses, we need larger and more diverse datasets to test if our findings hold for other social network and location-based service users. Finding or collecting such datasets, however, is non-trivial. We cannot rely on merely passively crawling an online social network [10], since we require information about those locations and events that were *not* shared as well as those that were. An open question is whether ESM-like studies are the only way to collect such information. While such studies provide a mechanism for reliably collecting high-quality data [5], the expense of the method means that the data collected in this way will be small and sparse. We are therefore interested in exploring new ways of collecting privacy preference data. In the meantime, we have made our dataset available through the CRAWDAD data archive [3], and encourage other data collectors to do the same.

7 Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council [grant number EP/G002606/1].

References

1. D. Anthony, T. Henderson, and D. Kotz. Privacy in location-aware computing environments. *IEEE Pervasive Comput.*, 6(4):64–72, Oct. 2007. doi:10.1109/MPRV.2007.83.
2. F. Ben Abdesslem, T. Henderson, S. Brostoff, and M. A. Sasse. Context-based personalised settings for mobile location sharing. In *Proc. RecSys PeMA Workshop*, Oct. 2011. Online at http://pema2011.cs.ucl.ac.uk/papers/pema2011_benabdesslem.pdf.
3. F. Ben Abdesslem, T. Henderson, and I. Parris. CRAWDAD data set st_andrews/locshare (v. 2011-10-12). Downloaded from http://crawdad.org/st_andrews/locshare, Oct. 2011.
4. F. Ben Abdesslem, I. Parris, and T. Henderson. Mobile experience sampling: Reaching the parts of Facebook other methods cannot reach. In *Proc. PUMP*, Sept. 2010. Online at http://scone.cs.st-andrews.ac.uk/pump2010/papers/benabdesslem.pdf.
5. F. Ben Abdesslem, I. Parris, and T. Henderson. Reliable online social network data collection. In A. Abraham and A. Ella Hassanien, editors, *Computational Social Networks: Mining and Visualization*. Springer-Verlag, London, UK, 2012. Accepted for publication, Online at http://www.cs.st-andrews.ac.uk/~tristan/pubs/sn2011.pdf.
6. A. R. Beresford and F. Stajano. Location privacy in pervasive computing. *IEEE Pervasive Comput.*, 2(1):46–55, Apr. 2003. doi:10.1109/MPRV.2003.1186725.
7. S. Bhagat, G. Cormode, B. Krishnamurthy, and D. Srivastava. Prediction promotes privacy in dynamic social networks. In *Proc. WOSN*, June 2010. Online at http://www.usenix.org/events/wosn10/tech/full_papers/Bhagat.pdf.
8. D. Corapi, O. Ray, A. Russo, A. Bandara, and E. Lupu. Learning rules from user behaviour. In *Proc. Workshop on Induction of Process Models*, pages 459–468, Sept. 2008. doi:10.1007/978-1-4419-0221-4_54.
9. S. J. Fusco, K. Michael, and M. G. Michael. Using a social informatics framework to study the effects of location-based social networking on relationships between people: A review of literature. In *Proc. ISTAS*, pages 157–171, June 2010. doi:10.1109/ISTAS.2010.5514641.
10. M. Gjoka, M. Kurant, C. T. Butts, and A. Markopoulou. Practical recommendations on crawling online social networks. *IEEE J. Sel. Areas Commun.*, 29(9):1872–1892, Oct. 2011. doi:10.1109/JSAC.2011.111011.
11. A. Gkoulalas-Divanis, P. Kalnis, and V. S. Verykios. Providing k-anonymity in location based services. *ACM SIGKDD Explor.*, 12(1):3–10, Nov. 2010. doi:10.1145/1882471.1882473.
12. M. A. Hall. Correlation-based feature selection for discrete and numeric class machine learning. In *Proc. ICML*, pages 359–366, June 2000.
13. B. Hecht, L. Hong, B. Suh, and E. H. Chi. Tweets from Justin Bieber's heart: the dynamics of the location field in user profiles. In *Proc. CHI*, pages 237–246, 2011. doi:10.1145/1978942.1978976.
14. B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady. Preserving privacy in GPS traces via uncertainty-aware path cloaking. In *Proc. CCS*, pages 161–171, 2007. doi:10.1145/1315245.1315266.

15. G. H. John and P. Langley. Estimating continuous distributions in Bayesian classifiers. In *Proc. UAI*, pages 338–345, Aug. 1995.
16. V. Kostakos, J. Venkatanathan, B. Reynolds, N. Sadeh, E. Toch, S. A. Shaikh, and S. Jones. Who's your best friend?: targeted privacy attacks in location-sharing social networks. In *Proc. UbiComp*, pages 177–186, Sept. 2011. doi:10.1145/2030112.2030138.
17. R. Larson and M. Csikszentmihalyi. The experience sampling method. In H. T. Reis, editor, *Naturalistic Approaches to Studying Social Interaction*, pages 41–56. Jossey-Bass, San Francisco, CA, USA, 1983.
18. Q. Li, J. Li, H. Wang, and A. Ginjala. Semantics-enhanced privacy recommendation for social networking sites. In *Proc. TrustCom*, pages 226–233, Nov. 2011. doi:10.1109/TrustCom.2011.31.
19. J. T. Meyerowitz and R. R. Choudhury. Realtime location privacy via mobility prediction: creating confusion at crossroads. In *Proc. HotMobile*, pages 1–6, Feb. 2009. doi:10.1145/1514411.1514413.
20. A. Narayanan, E. Shi, and B. I. P. Rubinstein. Link prediction by de-anonymization: How we won the Kaggle social network challenge. In *Proc. IJCNN*, pages 1825–1834, July 2011. doi:10.1109/IJCNN.2011.6033446.
21. J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Series in Machine Learning. Morgan Kaufmann, San Francisco, CA, USA, 1992.
22. J. J. Rodriguez, L. I. Kuncheva, and C. J. Alonso. Rotation forest: A new classifier ensemble method. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(10):1619–1630, Oct. 2006. doi:10.1109/TPAMI.2006.211.
23. A. Sadilek, H. Kautz, and J. P. Bigham. Finding your friends and following them to where you are. In *Proc. WSDM*, Feb. 2012. doi:10.1145/2124295.2124380.
24. E. Toch, J. Cranshaw, P. H. Drielsma, J. Springfield, P. G. Kelley, L. Cranor, J. Hong, and N. Sadeh. Locaccino: a privacy-centric location sharing application. In *Proc. UbiComp*, pages 381–382, 2010. doi:10.1145/1864431.1864446.
25. E. Toch, J. Cranshaw, P. H. Drielsma, J. Y. Tsai, P. G. Kelley, J. Springfield, L. Cranor, J. Hong, and N. Sadeh. Empirical models of privacy in location sharing. In *Proc. UbiComp*, Ubicomp '10, pages 129–138, Sept. 2010. doi:10.1145/1864349.1864364.
26. E. Toch, N. M. Sadeh, and J. Hong. Generating default privacy policies for online social networks. In *Proc. CHI Extended Abstracts*, pages 4243–4248, Apr. 2010. doi:10.1145/1753846.1754133.
27. A. S. Voulodimos, C. Z. Patrikakis, P. N. Karamolegkos, A. D. Doulamis, and E. S. Sardis. Employing clustering algorithms to create user groups for personalized context aware services provision. In *Proc. SBNMA*, pages 33–38, 2011. doi:10.1145/2072627.2072637.
28. I. H. Witten, E. Frank, and M. A. Hall. *Data mining : practical machine learning tools and techniques*. Morgan Kaufmann, San Francisco, CA, USA, 3rd edition, 2011.