Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media

Where Would You Go this Weekend? Time-Dependent Prediction of User Activity Using Social Network Data

Evgeniy Bart and **Rui Zhang** and **Muzammil Hussain**
Palo Alto Research Center
Palo Alto, CA 94394

Abstract

Predicting user activities and interests has many applications, for example, in contextual recommendations. Although the problem of predicting interests in general has been studied extensively, the problem of predicting when the users are likely to act on those interests has received considerably less attention. Such predictions of timing are extremely important when the application itself is time-sensitive (e. g., travel recommendations are irrelevant too far in advance and after reservations have already been made). Particularly important is the ability to predict likely future activities long in advance (as opposed to short-term prediction of imminent activities). In this paper we describe a comprehensive study that addresses this problem of making long-term time-dependent predictions of user interest. We have conducted this study on a large collection of visits to various venues of interest performed by users of Foursquare. We have built models that, given a user's history, can predict whether or not the user will visit a venue of a particular type on a given day. These models provide useful prediction accuracy of up to 75% for up to several weeks into the future. Our study explores and compares various feature sets and prediction methods. Of particular interest is the fact that venues interact with each other: to predict visits to one type of venue, it helps to use the history of visits to all venue types.

1 Introduction

Predicting user activities and interests has many applications, including targeted messaging, contextual recommender systems, transportation optimization, prediction of household energy use, and social coordination. General predictions of interest are often quite useful; for example, knowing that a user is interested in travel might enable offering that user travel-related discounts. However, predicting the timing of future activities, and predicting the time when a user's interest will be acted on, are much more valuable in some applications and crucial in others. For example, knowing that a user is likely to book plane tickets within the next week allows an advertiser to offer that user travel-related discounts that are much more timely and relevant during the booking period, and to avoid alienating the user with irrelevant offers the rest of the time. While many works in recommendation systems ensure relevance, we strive to guarantee timeliness. The latter is critical in the plane ticket booking scenario, because if we mis-time the discount offer and the user has committed to a regular ticket, he/she is very unlikely to take advantage of our discount later because the tickets are usually non-refundable.

Predictions of future activities can be short-term or long-term. Short-term prediction refers to predicting the "imminent" or "next activity" of a given user. While these predictions are of considerable interest in some applications (such as crowd management or transportation planning), they are unusable in many other applications. For example, recommendation systems cannot rely on short-term predictions because many human decisions (even trivial ones such as having dinner) are planned in advance. By the time a "short-term" recommendation becomes available, the user has already made a decision. Often, this decision is difficult to reverse. For example, airplane tickets are often non-refundable; and even changing a dinner venue at the last moment, especially when several people are involved, may be undesirable. Longer-term predictions are especially useful in such a context.

To our best knowledge, (Etter, Kafsi, and Kazemi 2012; Zhang et al. 2012) are the only previous attempts to address the problem of making time-dependent predictions. In this paper, we make a significant leap forward from these works, by presenting a method that can make both short-term and long-term activity predictions. We present a comprehensive study that considers a range of different algorithms and features / characteristics of the data. These considerations allow us to achieve greater accuracy against more thorough evaluations. We have used a large dataset of public check-ins by the users of the social networking site Foursquare. Each check-in corresponds to a visit by the user to a particular venue type (such as a restaurant, supermarket, train station, etc.). We have built models that, given a user's history, can predict whether or not the user will visit a venue of a particular type on a given day, up to several weeks in advance.

The remainder of this paper is organized as follows. In section 2, we briefly review previous approaches to the problem. In section 3, we describe our dataset in more detail. In section 4, we describe the models we have developed for prediction. The performance of these models is evaluated in section 5. We conclude in section 6.

2 Related work

Although the problem of predicting interests in general has been studied extensively (e. g., (Marlin 2003)), the problem of predicting when the users are likely to act on those interests has received considerably less attention.

In (Sadilek, Kautz, and Bigham 2012), the location of Twitter users is predicted based on friends whose location is known. However, that method is not applicable to the cases when friendship relationships are not available. An interesting direction for future research is to combine the method proposed here with the method in (Sadilek, Kautz, and Bigham 2012) for the case when both social relationships and personal visit history are available.

In (Scellato et al. 2011), sequence matching was used to make predictions of the user's likely next activity. In contrast, the method proposed here computes the descriptor of the user's history and makes predictions based on this descriptor. The two methods have complementary strengths and weaknesses. Sequence matching is potentially more precise; however, it also has greater potential to be disturbed by spurious events. In contrast, our descriptor-based method makes coarser-grained predictions, but is more tolerant of spurious events (which were a significant problem in our dataset). In addition, a descriptor can potentially incorporate much longer history than a literal sequence of all events. Up to a full week of history was used in some of our experiments (see section 5). An interesting direction for future research is combining the descriptor-based and the sequence-based methods.

The problem of making time-sensitive predictions has been addressed in (Etter, Kafsi, and Kazemi 2012; Zhang et al. 2012). However, both methods focused on predicting the "next" activity. This kind of predictors cannot be easily used to make long-term predictions. For example, suppose we would like to predict a user's Friday activities on Monday. Both of the above methods would require knowledge of all activities up to Thursday; this knowledge, however, is unavailable on Monday. The current paper extends the work in (Etter, Kafsi, and Kazemi 2012; Zhang et al. 2012) in several ways. First, we propose a method that allows making predictions up to several weeks into the future, with useful accuracy. Second, we define a much richer set of features that are used for prediction. This allows us to improve performance by 10% (3 percentage points) over (Zhang et al. 2012). Finally, we perform a thorough evaluation of our method on additional venue types (see section 5).

3 Dataset

Our dataset consists of over 180 million public check-ins by Foursquare users. When a user checks into a location with Foursquare, they are given the option to publicly broadcast that check-in via Twitter. We had a Twitter filtered search set up to listen for those tweets to collect check-ins.

The vast majority of users that publicly tweeted their check-ins only did so a few times. For our purposes, we selected a subset of 55,000 more active users with at least 200 check-ins each.

Each check-in event consists of a unique ID, an ID of the user who checked in and further information about the venue, including the full name and address, latitude, longitude, and a venue type. The venue type determines the category to which the venue belongs. There are three levels of granularity in Foursquare venue types. There are a total of eight tier-1 venue types. These are very general categories such as "food" or "art/entertainment". Each tier-1 venue type is further subdivided into several tier-2 venue types. For example, "art/entertainment" is further subdivided into "movie theaters", "museums", and several additional subtypes. Finally, each tier-2 venue type may be subdivided into several tier-3 venue types. For example, "museums" are subdivided into "art museums", "history museums", and others. There are 8 tier-1 venue types, 239 tier-2 venue types, and 348 tier-3 venue types.

4 Technical approach

We have focused on the task of predicting whether a given user will visit a particular venue type (called the 'target venue') on a given date (called the 'target date'). We have approached this task in a classical binary classification framework. For each user, for each venue type the user has visited, and for each date in the time period the user was active, a data point was constructed. The label of each data point was $+1$, or "visit", if the user visited the venue type on that date, and -1, or "no visit" otherwise. We have used a set of features, described below, to compute a descriptor for every data point. This descriptor contained information about the target date itself (such as the day of week for that date), as well as a concise summary of the user's history of visits up to that date (for example, the number of days since the last visit to a venue of the same type).

These data points were split into a training set and a test set. The split was performed in such a manner that for every user, all data points in the test set corresponded to later dates than the points in the training set. In other words, the (past) history of each user's check-ins was used to predict the future (unseen) check-ins. This setup ensured the testing performance corresponded well to what would be obtained in a realistic application (predicting the user's future check-ins based on the known history to date). In particular, one challenge in predicting future activities based on history is that users sometimes shift their activity patterns (Etter, Kafsi, and Kazemi 2012); this is accounted for in our testing setup.

A binary classifier was trained on the training set, and subsequently evaluated on the test set. For each classifier, the ROC (Receiver Operating Characteristic) curve was constructed, and the Equal-Error-Rate (EER) point on that curve was identified. EER is the point at which the false positive rate is equal to the false negative rate. These EER values were used to evaluate the performance of the classifiers. It is natural to use EER to make a general-purpose evaluation of a classifier's performance when the exact application is unknown. The reason is that the relative costs of false positives and false negatives may be different in different applications. EER provides a standardized way to evaluate performance when these costs are unknown. We have ongoing efforts that, when completed, will allow us to measure the application-dependent costs.

Feature sets

We have designed several types of features to perform prediction. Not all features were used for each prediction task. Here, we describe the entire feature set. In section 5, we describe which features were used for each task. The following features were investigated:

Day of week: the day of week for the target date. This feature helped capture the weekly cycle in the data, such as workdays/weekends, Friday night activities, etc.

Days since: an integer specifying the number of days since the last visit to the target venue type.

History: a set of binary (0/1) features. One feature for each venue type was used. Note that this included all venue types, not just the target venue type. Each feature indicated whether or not the venue type was visited on a particular day relative to the target day. Several options for choosing this 'history' day were evaluated:

- **today** The history of the user's visits on the target date, excluding visits to the target venue. These features helped capture the current user activities. The visits to the target venue were excluded to make the prediction task non-trivial.
- N **days ago** The history of the user's visits on the day N days before the target date. These features helped capture the recent history of visits. All venue types were used since excluding the target venue was no longer necessary.

5 Results

We have compared several types of classifiers on the prediction task outlined above. The classifiers we compared were logistic regression, Naive Bayes, SVM, mixture of logistic regressions and mixture of Naive Bayes models. No significant differences were observed between these classifiers; therefore, we only report logistic regression results unless indicated otherwise.

Venue interactions

In Table 1, we show the performance of different feature sets on the prediction task. The baseline feature set contains just two feature types, "days since" and "day of week". This baseline corresponds to the method used in (Zhang et al. 2012). The proposed method achieves useful prediction accuracy using just these features. However, note that the performance of the method improves by about 10% (3 percentage points) when features that describe visits to other venue types are added. This improvement indicates that the visits to different venues are not independent, but rather interact in various ways.

Feature set	tier-1	tier-2	tier-3
Target venue	28%	34%	34%
Target venue + Other venues	25%	33%	33%

Table 1: Prediction performance with various feature sets. Equal error rates are shown. Lower numbers indicate better performance. The "target venue" feature set (first row) contains just two feature types, "days since" and "day of week". The feature set in the second row uses features that describe visits to other venue types in addition to the target venue features. We have tried predicting tier-1, tier-2, and tier-3 venue type check-ins. The EER for each task is shown in the three rightmost columns. (Corresponding accuracy values are shown in Table 2.) As can be seen, using information from other venue types improves performance compared to using only the information related to the target venue. This indicates that the activities of the users are not independent, but rather interact in various ways.

Feature set	tier-1	tier-2	tier-3
Target venue	72%	66%	66%
Target venue + Other venues	75%	67%	67%

Table 2: Accuracies at the EER point for various feature sets. The format is the same as in Table 1.

An interesting direction for future research is to describe the ways in which these interactions occur. Of particular interest is the idea that a higher-level concept may underlie several related check-ins. For example, check-ins at a transportation venue, an office, a restaurant, and another transportation venue might correspond to a concept of "a workday in an office", while several check-ins at bars and restaurants on a Friday night might correspond to a concept of a "pub crawl". Automatically discovering such higher-level concepts is a subject of future research.

The accuracy at the EER point is, of course, 100% – EER. These accuracy values are shown in Table 2. It is desirable to report accuracy values at the EER point, especially for unbalanced datasets. For example, for tier-3 venues, only 2–8% of the days included any visits. As a result, a method that always predicts "no visit" will achieve over 90% overall accuracy. This number is inconvenient to work with, however, because it is composed of 100% accuracy on negative examples and 0% accuracy on positive examples. In contrast, X% accuracy at EER always corresponds to X% correct on positive examples and X% correct on negative examples.

It is also of interest to compare the performance to the method used in (Zhang et al. 2012). However, the accuracy results in (Zhang et al. 2012) were not reported at the EER point; therefore, they are not directly comparable to the results in Table 2. We have computed the accuracy of our method at the ROC point used in (Zhang et al. 2012). Our accuracy is 87% correct, compared to 76% correct for (Zhang et al. 2012)).

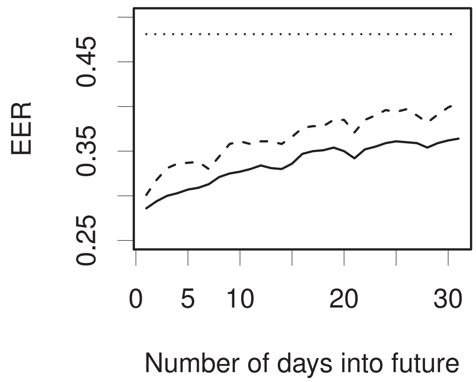

Figure 1: Prediction performance as function of the time horizon. Y axis: equal error rate (lower numbers indicate better performance). X axis: the number of days into the future the prediction is being made. Dashed line: prediction using "day of week" and N-day old history features. Solid line: prediction using "day of week", N-day old history features, and N-day old history features for the same day of week as target date. Dotted horizontal line: EER with only the "day of week" feature; this can be thought of as asymptotic EER of predicting infinitely far into the future (when history features become irrelevant).

Predicting future activities

The performance figures reported in Table 1 describe short-term prediction performance (i. e., predicting the action today based on recent history). Predicting activities further out in the future is of considerable interest in many applications (such as optimizing the timing of discount offers). The method proposed here can be readily applied to this problem.

The only necessary modification is omitting the feature "Days since", because its value is unknown when making predictions for any day other than today. For example, if on February 11th we'd like to make a prediction for visits on February 15th, computing the value of "Days since" for February 15th requires knowing whether there was a visit on any of the dates from 12th through 14th; these, however, are unknown on February 11th.

We have applied the method proposed here to predicting user visits N days into the future. We have used the "Day of week" feature (since it can be computed for an arbitrary future date), as well as the "history" features for N days earlier than the target date.

The performance of the method is shown in Figure 1 (dashed line). As expected, the performance decreases when trying to predict events further in the future, but remains useful for up to several weeks. Note also the strong weekly seasonality: the error rate drops considerably when making predictions an integer number of weeks into the future. The reason is that users' activities exhibit strong week-based seasonality, and it is advantageous to use past data from the same day of week in addition to the more recent data. The solid line in Figure 1 shows the prediction performance based on two types of history features: the features for N days earlier than the target date (same features as those used for the dashed line), as well as history features for the latest date, older than N days, that is the same day of week as the target date. Note again that both types of history features are at least N days old and can therefore be used for predicting N days into the future. As can be seen, this combination improves performance considerably over the previous method.

6 Conclusions

We have presented a method for predicting the timing of user activity. The motivation behind the method is that while knowing the general set of interests of a customer is useful, knowing the time when the custormer is likely to act on those interests is even more valuable in a variety of applications. The method we presented was evaluated on a large database of Foursquare check-ins. We have shown that the proposed method can achieve useful accuracy (up to 75% correct at EER) and predict the timing of activities up to several weeks into the future. An intriguing observation is that check-ins to different venue types are not independent, but rather interact in various ways. Exploiting these interactions allows the method to significantly improve the prediction accuracy.

References

Etter, V.; Kafsi, M.; and Kazemi, E. 2012. Been there, done that: What your mobility traces reveal about your behavior. In *Mobile Data Challenge 2012 Workshop at Pervasive*.

Marlin, B. 2003. Modeling user rating profiles for collaborative filtering. In *NIPS*.

Sadilek, A.; Kautz, H.; and Bigham, J. P. 2012. Finding your friends and following them to where you are. In *Proceedings of the fifth ACM international conference on Web search and data mining*, 723–732.

Scellato, S.; Musolesi, M.; Mascolo, C.; Latora, V.; and Campbell, A. T. 2011. NextPlace: A spatio-temporal prediction framework for pervasive systems. In *Pervasive*.

Zhang, R.; Price, B.; Chu, M.; and Walendowski, A. 2012. Location-based predictions for personalized contextual services using social network data. In *AwareCast workshop in conjunction with Pervasive 2012*.