Research Article

Twitter Can Predict Your Next Place of Visit

Arun Chauhan, 1 Ravi Tejwani, 2 and Durga Toshniwal 1

1 Computer Science, Indian Institute of Technology, Roorkee 247667, India
2 IBM Watson, Cambridge, MA 02142, USA

Correspondence should be addressed to Arun Chauhan; aruntakhur@gmail.com

Received 25 September 2015; Revised 19 January 2016; Accepted 26 January 2016

Academic Editor: Antonio J. Jara

Copyright © 2016 Arun Chauhan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The present work focuses on predicting users’ next place of visit using their past tweets. We hypothesize that tweets of the person have predictive power on his location and therefore can be used to predict his next place of visit. This problem is important for location based advertising and recommender based services. To predict the next place of visit, we calculate the probabilities of visiting different types of places using bank of binary classifiers and Markov models. More specifically, we train bank of binary classifiers on past tweets and calculated the probabilities of visiting next places. Since bank of binary classifiers is based on a bag-of-words model, to account for time of last visited place and place itself, we built Markov models for different time duration to calculate probabilities of visiting next place. Empirical evaluations show that by combining the probabilities obtained from bank of binary classifiers and Markov models the accuracy of predicting next place increased from 65% to 80%.

1. Introduction

How the life would be if we know the location of desired and intended place around us anywhere in the world. Though it seems to be superficial, the development in wireless and location acquisition technology helps us to build systems that solve this problem to some extent. Because of location acquisition technology, nowadays, it is not difficult to find out the places in given proximity. The bigger challenge is in predicting places for user according to his need and intent in given proximity. The existence of microblogging services like Twitter and Facebook provides a means to know the person’s values and needs in better way.

People share their thoughts and happening on Twitter along with mundane information. In recent few years, the popularity of Twitter is growing exponentially. As of now, at the time of writing this paper the number of users on Twitter are 300 million producing 500 million (https://about.twitter.com/company) tweets per day. With mundane information, people also share their activities like what they are planning, visited places, experiences of visited places, with whom they are going, how they are feeling, and so forth. Availability of such valuable information motivates us to build a system that can predict the future places of visits for the user according to their intent and behavior.

In this paper, we hypothesize that future places of visit can be predicted considering two important factors that are (i) previous visited place and (ii) recent tweets. What people write on their timeline reflects their intent and need, which have a significant role in deciding the next activity to be performed. Similarly, previous visited place also has a major role in deciding the next place to be visited. For example, after having lunch people have coffee. Based on given hypothesis we propose a novel approach for predicting next places of visit. We use bank of binary classifiers (BBC) for predicting the probability of visiting next place using recent tweets only. And to account for time of last visited place and place itself, we build Markov models for different time duration to calculate probabilities of visiting next place. Empirical evaluation shows that by combining the probabilities obtained from bank of binary classifiers and Markov models the accuracy of predicting next place increased from 65% to 80%.

(a) assigning location to tweets if relevant information is


presents; (b) extracting features from past tweets that are used to train the models; (c) building models using bank of binary classifiers; (d) using contextual information to enhance the accuracy of predicting the next place. Here contextual information is time of previous visited place and place itself.

We crawl more than 4600 Twitter timelines for exhaustive experimentation. Ground truths are extracted from these timelines using Google Places API and by analyzing tweets for mentions of visited places. We extract features from users timelines by using our proposed algorithm for finding out the optimum length of past tweets for prediction. We build models and do performance analysis. Our best model yields 80% of accuracy for top 5 predictions. We evaluate our approach on new users also and show that performances of models are similar to the seen users to model. The major contribution of this work is as follows:

1. Building generic models for predicting future places of visit using recent tweets only.
2. Building Markov models for predicting next place to be visited given previous place visited and time since it is visited.
3. Ensemble of both models to enhance the model accuracy in predicting next place.

The remainder of the paper is organized as follows. Section 2 presents the proposed approach in detail for predicting next place of visit. Next, Section 3 discusses the experiments and analysis of results. In Section 4 we review relevant literature work. Finally, Section 5 concludes the paper.

2. Proposed Approach

There are two points we want to mention before explaining the 4 steps of proposed approach. Firstly, we build the generic prediction model that captures the relationship between vocabulary used in past tweets and visited places, without considering user’s demographics. The main advantage of this approach is that we do not need individual specific training data for predicting their next location.

Secondly, our focus would be the category of the establishment like restaurant, supermarket, pub, and gym, rather than the specific establishment. By doing this, both establishment owners and users can get mutual benefits. For example, if proposed system is predicting restaurant in given spatial proximity, then all the owners of restaurants in the proximity of user’s location can approach user with their available offers for promotions. Along with this, the user can also have the option to choose place according to his own suitable interest.

Four steps of our proposed approach are as follows.

2.1. Assigning Location to Tweets. From here onwards we are using location and place of visit interchangeably. Assigning location to tweet is two-step process. First, filter all those tweets having coordinates and location information. For location information, we use a regular expression ("I’m at" or "@"'). Then by using Google Places API [1] (GPA), get all the places around geocoordinates of the given tweet.

| Past Tweets | Categories visited |
|-------------|--------------------|
| P1          | C₁, C₃             |
| P2          | C₁, C₃, C₄         |
| P3          | C₄                 |
| P4          | C₃, C₅             |
| P5          | C₁, C₄             |

If place name present in a tweet is present among places returned by GPA, then that tweet is labeled by the name and categories of location. We want to mention that GPA also returns categories of places which may be more than one. For matching the place name, we extract three words after regular expression from a tweet and find out the ratio r, with total words in each place name returned by GPA. If $r > 0.75$, we label the tweet with the first match among GPA places.

2.2. Extracted Features. For predicting next place of visit, we use past tweets to infer the intent and interest of the user. We analyze from the ground truths recovered from timelines that people post activity related text before performing that activity. For example, user tweeted, “We are planning to go for some fun” before visiting the central park in New York. The time interval of activity related post may vary from weeks to hours based on the activity to be performed. For example, user interested in cricket match going to happen next month has already started tweeting about the event whereas a user interested in going to a restaurant in the evening tweets just a few hours before visiting restaurant. Therefore, we built independent binary classifiers for each category using appropriate size of the window of past tweets. Here window stands for the time window of past tweets that is to be taken to form feature vectors. The window size is found empirically for each classifier/category independently and explained in next section.

To form feature vectors, first we label past tweets (concatenation of past tweets in given window) with the categories of location, visited by the user, just after posting these past tweets. Note that the location which user has visited can have more than one category, and, therefore, the labels of past tweets may be more than one. It is worth mentioning here that while concatenating past tweets, we are not considering location tweet (tweet that has location information). To build a binary classifier for category $C_i$ corresponding data set $D_i$ is formed. To form $D_i$, those past tweets which are labeled by $C_i$ are treated as positive samples and remaining as negative samples. Hence for training 100 (total number of categories) binary classifiers, we have 100 corresponding data sets. For example, toy data set in Table 1 is having four different categories. Table 2 shows four binary data sets ($D_i$) for four categories, respectively, as explained above.

2.3. Bank of Binary Classifiers (BBC) for Predicting Future Location. For each category, we build independent binary classifier and the size of the window of past tweets for each
Table 2: Four binary datasets constructed for each category respectively using dataset given in Table 1. Here - denotes the negative samples.

| (a) Data Set $D_1$ for category $C_1$ |
|---|
| Past Tweets | Categories Visited |
| P1 | $-C_1$ |
| P2 | $C_1$ |
| P3 | $-C_1$ |
| P4 | $-C_1$ |
| P5 | $C_1$ |

| (b) Data Set $D_2$ for category $C_2$ |
|---|
| Past Tweets | Categories Visited |
| P1 | $C_2$ |
| P2 | $-C_2$ |
| P3 | $-C_2$ |
| P4 | $C_2$ |
| P5 | $-C_2$ |

| (c) Data Set $D_3$ for category $C_3$ |
|---|
| Past Tweets | Categories Visited |
| P1 | $C_3$ |
| P2 | $C_3$ |
| P3 | $-C_3$ |
| P4 | $C_3$ |
| P5 | $-C_3$ |

| (d) Data Set $D_4$ for category $C_4$ |
|---|
| Past Tweets | Categories Visited |
| P1 | $-C_4$ |
| P2 | $C_4$ |
| P3 | $C_4$ |
| P4 | $-C_4$ |
| P5 | $C_4$ |

A classifier is determined empirically. The steps for determining the window size of past tweets for category $C_i$ are given in Algorithm 1.

In Algorithm 1, for each category, we vary window size $w$ (in hours) from 6 hours to 600 hours to construct feature vectors. Then, for each window size $w$, we form feature vectors that are used to train the binary classifier. Finally, the window size of a particular category is set, based on the best performance of classifier among all window sizes. For constructing binary data set of each category, function getBinaryDataSet($D$, $C_i$, $w$) is called with parameters that are timelines of all users ($D$), name of category ($C_i$), and size of window of past tweets in hours ($w$). For all location tweets available on all timelines in $D$, we form feature vectors by concatenating past tweets in $w$ hours before location tweet (excluding location tweet) and labeled these feature vectors with categories of the location visited by the user in location tweet. Those feature vectors having label $C_i$ are denoted by the “positive” value of the class attribute in the binary data set and remaining feature vectors by “negative” value of the class attribute, as shown in Table 4. Hence, this function returns the binary data set $D_i$ for category $C_i$. Then binary model $H_i$ is built using returned data set $D_i$ for category $C_i$ and accuracy of models are calculated on validation data set to perceive the performance of model on given window size ($w$). Similarly, model performance is evaluated on different window sizes and based on the best performance of model, respective window size of past tweets for that category is set. This optimum window size of category $C_i$ is denoted by $W_i$. Once the window size of each category is determined, we use the respective window size in constructing feature vectors for that category. These feature vectors are used in training and prediction.

Figure 1 shows the block diagram of prediction of places of visit using past tweets. From past tweets we form feature vector $FV(C_i)$ using window size ($W_i$) for $i$th category determined in Algorithm 1. Then $FV(C_i)$ is used to infer the probability of visiting the category $C_i$. Please note that while predicting the probability of visiting category $C_i$, we use past tweets depending on $W_i$ for $H_i$ classifier at prediction time, as it is hypothesized that history of activity related tweets depends on activity to be performed. Therefore, this probability is denoted by $P(C_i | W_i)$. For our purpose, we use Naive Bayes as binary classifier $H_i$ in training and testing.

2.4. Considering Previous Visited Place and Time. We consider two other important factors in predicting next place of visit that are previously visited place and amount of time before it is visited (i.e., time duration) along with the recent tweets. For example, predicting next place of visit as restaurant, after user has visited restaurant only, within an hour, is not considered as a good prediction. For our purpose, to track the duration of time between visits of next and previous place, we discretize the time into one-hour slots. We capture the human behavior of visit in a very simplified manner by building Markov chain for different time duration. Therefore, for each time duration, we form Markov chain and transition matrix. In transition matrix $t_h$, entry $t_h(i, j)$ represents the probability of visiting place $j$ after place $i$ between time interval $[h, h + 1)$ hours. Let $n_h(i, j)$ be the number of times users’ visited place $j$ after place $i$ between time interval $[h, h + 1)$ hours; then

$$t_h(i, j) = P_h(C_j | C_i) = \frac{n_h(i, j)}{\sum_k n_h(i, k)}.$$

We form transition matrix for each time duration (hours) in set $\{1, 3, 6, 12, 24\}$. The probability of visiting next category $C_i$ after $h$ hours given previous visited location having categories $C_{a_1}C_{a_2}\cdots C_{a_b}C_{a_b+1}\cdots C_{a_b}$ ($C_i$ is name of category whereas $C_{a_b}$ is variable that can be any category. E.g., $C_{a_1}C_{a_2}$ can be $C_{a_2}C_{a_3}$) is calculated as follows. Recall that visited location may have more than one category:

$$P_h(C_i | C_{a_b}) = P_h(C_{a_b+1} | C_{a_b}) P_h(C_{a_b}) = P_h(C_{i} | C_{a_b}).$$
**Input:** $D$: Timelines of all users after labeling, as described in Section 2.1.

**Output:** Window size of each $i$th category that is $W_i$.

1. **foreach** category $C_i$ in $D$ **do**
   2. $W_i = 6$ hours;
   3. $Acc_i = 0$;
   4. $w = 0$;
   5. **repeat**
      6. $w = w + 6$;
      7. $D_i = \text{getBinaryDataSet}(D, C_i, w)$;
      8. train binary classifier $H_i$ using data set $D_i$;
      9. calculate accuracy $A_i$, of classifier $H_i$ on validation data set;
     10. **if** $A_i > Acc_i$ **then**
         11. $Acc_i = A_i$;
         12. $W_i = w$;
     13. **end**
   14. **until** $w \leq 600$;
15. **end**

$\text{getBinaryDataSet}(D, C_i, w)$:
1. For each location tweet in $D$, concatenate past tweets in $w$ hours, excluding location tweet;
2. Concatenated tweets are feature vectors labeled by the categories of location;
3. Feature vectors labeled with $C_i$ are considered as positive samples and the rest as negative samples;
4. **return** $D_i$

**Algorithm 1:** For determining the window size of past tweets use to construct feature vectors of each category.

![Diagram](image.png)

**Figure 1:** Predicting probability of visiting category $C_i$ given past tweets in window size $W_i$.

Considering independence assumption between previous visited categories we can write (2) as

$$P_h(C_i \mid C_{a\ldots x}) = \frac{P_h(C_a \mid C_i) P_h(C_b \mid C_i) \cdots P_h(C_x \mid C_i) P(C_i)}{P(C_a) P(C_b) \cdots P(C_x)}$$

$$= \frac{(P_h(C_i \mid C_a) P(C_a) / P(C_i)) (P_h(C_i \mid C_b) P(C_b) / P(C_i)) \cdots (P_h(C_i \mid C_x) P(C_x) / P(C_i)) P(C_i)}{P(C_a) P(C_b) \cdots P(C_x)}.$$ 

(3)
After simplifying, we can write (4) as
\[
P_h(C_i | C_{ab-x}) = \frac{P_h(C_i | C_a) P_h(C_i | C_x) \cdots P_h(C_i | C_x)}{P(C_i)^{(ab-x)|-1}},
\]
where \(P_h(C_i | C_x)\) are conditional probabilities estimated in (1) and \(P(C_i)\) is estimated as follows:
\[
P(C_i) = \frac{n(C_i)}{\sum_k n(C_k)}.
\]
(6)
n\((C_i)\) represents the number of times user visited the category \(C_i\) in the training set. Equation (5) uses information of both factors that are previously visited place and how much time before it is visited by picking the appropriate transition matrix. We refer to this Markov modeling which considers time information also as MMs for the sake of brevity.

In order to predict the next place of visit using past tweets and amount of time before which categories are visited by user, we combine these two probabilities that are \(P(C_i | C_{ab-x})\) (estimated in (5)) and \(P(C_i | W_i)\) (refer to Figure 1) under the assumption of independence. We refer to the combined classifier \(P(C_i | W_i, C_{ab-x})\) as CC for the sake of brevity. Consider
\[
P(C_i | W_i, C_{ab-x}) = \frac{P(W_i, C_{ab-x} | C_i) P(C_i)}{P(W_i, C_{ab-x})}.
\]
(7)
As \(W_i\) and \(C_{ab-x}\) are assumed to be independent, therefore we can write (7) as follows:
\[
P(C_i | W_i, C_{ab-x}) = \frac{P(W_i | C_i) P(C_i | C_{ab-x})}{P(C_i) P(W_i | C_{ab-x})}.
\]
(8)
\[
P(C_i | W_i, C_{ab-x}) = \frac{P(C_i | W_i) P(W_i)}{P(C_i) P(C_i | C_{ab-x}) / P(C_i) P(W_i) P(W_i | C_{ab-x})}.
\]
(9)

### 3. Experiments

We have done exhaustive experiments on the data set crawled from Twitter using publicly available Twitter API. We have collected 4606 users' timelines that have tweeted at least once from New York between April 24, 2014, and April 29, 2014. Each timeline contains approximately 3200 recent tweets. Also, the tweet rate of these users is at least 20 tweets per day. Among 4606 users the number of location tweets on timelines is highly variable. Though we have started from New York the locations visited by users in our data set are around the world.

#### 3.1. Data Sets

For evaluation, we divide our data set into two subsets according to the number of locations on user's timeline. Therefore, among 4606 users, those who have more than 60 location tweets (tweet that has location information) on their Twitter timeline are considered in the first subset named Data Set 1 and the remaining users in the second subset named Data Set 2. Table 3 has shown both data sets mentioned.

**Training Data Set (TDS).** From Data Set 1, we use oldest \(n - 50\) location tweets of each user for training the binary models, where \(n\) is total number of location tweets available on their timeline. Tweets on timelines are ordered in reverse chronological order.

**Testing Data Set.** For evaluating the performance of models on both seen and unseen users, we use two different sets for testing, named as Test Set 1 and Test Set 2.

(a) Test Set 1 (TS1): we use the remaining latest 50 location tweets of each user from Data Set 1 to form feature vectors.

(b) Test Set 2 (TS2): we use all location tweets from Data Set 2 to form feature vectors.

#### 3.2. Evaluation

First we present methods to compute the accuracies of proposed models one by one in the following subsections and then propose few baselines for comparing our proposed models.

3.2.1. Using Markov Models (MMs) Only. We form transition matrix \(t_h\) from training data set, where \(h \in \{1, 3, 6, 12, 24\}\) in hours, as explained in (1). By using the ground truth of test sets, we compute the accuracy of model as follows:

(1) For every test instance, first we find out both time duration and categories the user has visited just before prediction. Here test instance is the location of a tweet which we want to predict.

(2) Duration is then discretized to integer value \(h\), such that if duration lies within time interval \([h, h+1]\) then transition matrix \(t_h\) is used for computation in step (3).

(3) If previously visited categories are \(C_{ab-x}\), then by using \(t_h\) in (5) we estimate the probability of visiting each category present in training data set.
Table 3: Division of users according to number of location tweets on their timelines.

| Data Set        | Number of Users | Number of Location |
|-----------------|-----------------|--------------------|
| Data Set 1      | 1706            | >60                |
| Data Set 2      | 2900            | ≤60                |

Table 4: Description of training and testing data sets.

| Data Set                  | # Samples | Detail                                      |
|---------------------------|-----------|---------------------------------------------|
| Training Data Set (TDS)   | 260,423   | Oldest n – 50 location of each user from Data Set 1 |
| Testing Data Set 1 (TS1)  | 84,470    | Latest 50 location of each user from Data Set 1 |
| Testing Data Set 2 (TS2)  | 56,774    | All location of each user from Data Set 2    |

(4) Considering top N categories in nonascending order in the previous step (3), if we recover the ground truth, then we take this test instance as correctly classified. Hence, the accuracy of the model can be defined as follows:

\[
\text{Acc@N} = \frac{M}{K},
\]

where M represent the number of test instances accurately classified and K represents total number of test instances.

3.2.2. Using Bank of Binary Classifiers (BBC). By using the ground truth of test sets, we compute the accuracy of model as follows:

(1) For a given test instance (location tweet) we formed feature vector \(FV(C_i)\) for each category \(C_i\) as described in Section 2.3. We compute the probability \(P(C_i|W)\) of visiting the category \(C_i\) by using the binary classifier \(H_i\) modeled in the training phase.

(2) By considering top N categories according to probabilities predicted by binary classifiers, if we recover the ground truth, then we take this test instance as correctly classified. The accuracy of this proposed classifier is calculated in the same way as mentioned in (11).

3.2.3. Using \((P(C|W,C_{ab=x}))\) (CC). The evaluation method is similar to evaluation method illustrated for BBC. For every test instance, we compute the probability of visiting each category by using (10). Considering top N categories in nonascending order, if we recover the ground truth, then we classify this test instance as correctly classified. Accuracies of this model are also calculated in similar fashion as mentioned earlier in (11).

3.2.4. Baselines. Our proposed approach is generic and can be applied to both seen and unseen (new) users. As per our knowledge, there is no such literature available that predicts the future place of visit based on only recently used words and latest visited location without using user’s demographics. Hence, for evaluating the performance of models, we proposed the following baselines.

Baseline Model 1 (BM1). In this baseline, we have considered the most frequent check-in category in the training data. Let \(n(C_i)\) be the number of times users visited the category \(C_i\) in the training set. Then,

\[
P(C_i) = \frac{n(C_i)}{\sum_j n(C_j)},
\]

where \(n(C_i)\) represent the number of times user has visited the category \(C_i\).

Baseline Model 2 (BM2). Markov models are built for predicting the next place of visit based on the latest visited location, say, \(C_i\), by the user. The probability of visiting \(C_i\) is estimated by computing a fraction of number of user’s visits to \(C_i\) after \(C_L\) in total number of visits to any category after \(C_L\). For example, the probability of visiting category \(C_i\) is estimated as follows:

\[
P(C_i) = \frac{n(C_i,C_L)}{\sum_j n(C_L,C_j)},
\]

where \(n(C_i,C_L)\) represents the count of visiting categories \(C_i\) after visiting \(C_L\) in order.

Here, we want to mention that, in Section 2.4, we form Markov models (MMs) considering amount of time between consecutive visits. Therefore, if we use this baseline model to predict future place, we have only one transition matrix, but in MMs, we have different transition matrix depending upon the granularity of time.

Baseline Model 3 (BM3). As described in Section 2.3, we discuss the need of different window size for each category while predicting the future place of visit. This baseline is defined to show the importance of our approach for deriving window size for each category. For this, we use fixed window size for each category and show the performance against our approach used in Algorithm 1. Similar to BBC in Section 2.3, in this baseline, Naive Bayes is used as binary classifier \(H_i\) (see Figure 1), for training and testing. We use five different window sizes that are 5, 10, 20, 30, and 100 hours.

3.3. Results. We conduct two sets of experiments. In the first set, we compare the performance of model BBC proposed in Section 2.3 with Baseline Model 3 (BM3). In the second set of experiments, we compare the performances of proposed models with Baseline Models 1 and 2.

3.3.1. \((P(C|W))\) versus BM3. The objective of this set of experiments is to show that BBC performs better when feature vectors are constructed by using window sizes derived by Algorithm 1 for training and testing. We want to recall that window size derived for category \(C_i\) may not be equal to category \(C_j\), where \(i\) and \(j\) are some arbitrary category, as window size depends upon the performance of the binary classifier.
For this, first we show the performance of the classifier on same window size for each category (BM3). One by one, we use five different window sizes that are 5, 10, 20, 30, and 100 hours. Results are shown in Figure 2. \( \text{Win}_N \) represents the performance of BM3 when window size of past tweets is set to \( N \) hours. From these results, we observe that BM3 performs better when the size of a window is 10 hours for each category. Also, BM3 performs similarly on both test sets, that is, seen (TS1) and unseen (TS2), which validates that this classifier can be applied to a wide variety of users. These results also support our hypothesis that words in tweets posted by users are highly correlated with the future activity or the location of that activity done by that user. By learning the relation of words and locations, BM3 infers the future location using words with high accuracy.

Results shown in Figure 2 validate the existence of the relationship between words and future location. By observing the text in a window, we found out that using same window (BM3) for all categories does not model the problem appropriately. If the window size is 5 hours then the words are very few which results in erroneous predictions. But if we increase the window size, then some words may come which have no role in predicting the current location. For example, tweet like “now looking for some fun” tweeted 30 hours before has no significance in deciding the current activity. From the above results, we found optimum window size is 10 hours for BM3.

For appropriate modeling, as discussed in Section 2.2, we derive window size for each category by using Algorithm 1 for training and testing. We compare these two approaches that are BM3 (10 hours) and \( P(C \mid W) \). Results in Figure 3 show that we can enhance the accuracy of prediction by considering appropriate duration of tweets’ window.

3.3.2. Comparison of Proposed Models with BM1 and BM2.

In this section, we compare the proposed models with BM1 and BM2. The performance of our proposed models outperforms both baselines as shown in Figure 4. We can see that MMs perform much better than BM2. The reason of this improvement is that when we use BM2 only, we have only one transition matrix with no time information in it. Thus, the only input to BM2 is previous visited category. But when we use MMs for prediction, we have two input parameters that are previous visited category and the time duration when it is
visited. Based on the time duration we pick the corresponding transition matrix for computing predictions. The information of visiting preference according to time is lost in BM2, hence resulting in erroneous predictions. $P(C \mid W)$ is explained in Section 3.3.1 in detail. From accuracies given by MMs and $P(C \mid W)$, we can infer that both recent tweets and previous visited category with duration have almost equal role in predicting next place of visit. But when both these model accuracies are combined under the independent assumption, the prediction rate improves significantly as seen in Figure 4. From this improved accuracy we can infer that the accuracy of proposed model can be enhanced by having more contextual information of the user which helps the model in understanding the user in better way.

4. Related Work

In last decade, researchers have used Twitter data for predicting various things in future. Bollen et al. [2] used tweets for predicting stock market index with high accuracy. Similarly Asur and Huberman [3] predict box-office revenues of movies in advance using related tweets.

Tweets are also explored heavily for inferring users interests for commercial purposes. References [4–6] have used different techniques over tweets for inferring user interest and show that tweets contain a lot of valuable information related to user interest.

For recommendation also people used Twitter data. For example, Sadilek et al. [7] used tweets to recommend those restaurants that user should not go. References [8–10] proposed a recommender system for news recommendations by modeling the user profile and exploiting the tweet-news relationship.

With the exponential growth in usage of smart phones, now users publish millions of tweets frequently from anywhere and at any time. Because of this, one other field emerged that studies the mobility prediction of users; that is, where the user will be? or where he was when given tweets were published? References [11–13] have shown that by using Twitter data we can predict the location of user with high accuracy. But the granularity level of predicting the user location using tweets is at either the country level or regional level. For example, Han et al. [14] predict user location, that is, country or region by identifying location indicative words (i.e., frequent word used at the specific location), in contrast to our approach where we predict locations such as shop, church, and restaurant, where granularity level is very deep.

Some of the efforts have been made for predicting user location such as restaurant and shops, but the data used there are generated by Location Based Social Networks (LBSN). Using LBSN for prediction is very different from using Twitter, as Twitter data is very unstructured and challenging in comparison to structured data of LBSN. References [15–19] have explored LBSN (Foursquare) for recommending or predicting next location to the user.

Some people used text published by users for deriving personality traits and based on common traits recommendations have been made. For deriving personality traits Linguist Inquiry and Word Count (LIWC) [20] had been used frequently. References [20–29] have shown that lexicons used by people can be used for understanding their personal values and how to use these traits for a recommendation. Though all these approaches have been used extensively in analyzing personality traits, these also have shortcomings of predefined word category correlation. Alternatively, Schwartz et al. [30] used rather different approach for using vocabulary known as open-vocabulary technique (using all set of words available on social media) in comparison to closed-vocabulary technique, Linguistic Inquiry and Word Count (LIWC) [20], where some predefined sets of words are used for deriving the personality traits. In their study, they have shown that by using the open-vocabulary approach they got higher state-of-the-art accuracy in predicting gender in comparison to LIWC by exploring latent factors that are not captured by closed-vocabulary approach. Motivated by this approach, we have used all words available on timelines for modeling the users’ behavior.

5. Conclusion

In the present work, we study the problem of predicting next place of visit using tweets. We proposed a methodology for
predicting next place of visit for user. For experiments, we crawled more than 4600 users’ timelines from Twitter. For modeling and generating ground truths, we labeled those tweets with visited locations where relevant information is present. For labeling tweets, we propose simple pattern matching technique labeling tweets with high accuracy. We have also proposed an algorithm for deriving optimal length of window for each category for deriving feature vectors which is used in training and testing of BBC (bank of binary classifiers). From this trained model, BBC (Naïve Bayes), we compute the probabilities of visiting next place. To account for the time of last places visited, we trained Markov models that also compute probabilities of visiting next places. Assuming the independence of probabilities produced by bank of binary classifiers and Markov models, we combined these two probabilities. From experiments, we found out that probabilities of visiting next place increased up to 80% from 65%. This shows that our model can be potentially used in location based advertising, intelligent resource allocation, and so forth. Also, similar performance of proposed model in location based advertising, intelligent resource allocation, and so forth. Also, similar performance of proposed model on both seen and unseen datasets make it applicable to wide variety of users.

**Conflict of Interests**

The authors declare that there is no conflict of interests regarding the publication of this paper.

**References**

[1] "Google Places API," https://developers.google.com/places.

[2] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, 2011.

[3] S. Asur and B. A. Huberman, "Predicting the future with social media," in *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence (WI’10)*, pp. 492–499, Toronto, Canada, September 2010.

[4] P. Bhattacharya, M. B. Zafar, N. Ganguly, S. Ghosh, and K. P. Gummadi, "Inferring user interests in the twitter social network," in *Proceedings of the 8th ACM Conference on Recommender Systems (RecSys ’14)*, pp. 357–360, Foster City, Calif, USA, October 2014.

[5] C. Budak, A. Kannan, R. Agrawal, and J. Pedersen, "Inferring user interests from microblogs," Tech. Rep., 2014.

[6] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Inferring user interests from tweet times," in *Proceedings of the 1st ACM Conference on Online Social Networks (COSN ’13)*, pp. 235–240, ACM, Boston, Mass, USA, October 2013.

[7] A. Sadilek, S. P. Brennan, H. A. Kautz, and V. Silenzio, "nEmesis: which restaurants should you avoid today?" in *Proceedings of the 1st AAAI Conference on Human Computation and Crowdsourcing (HCOMP ’13)*, AAAI, 2013.

[8] N. Jonnalagedda and S. Gauch, "Personalized news recommendation using twitter," in *Proceedings of the IEEE International Conference on Web Intelligence and Intelligent Agent Technology (WIC/ACM ’13)*, pp. 21–25, Atlanta, Ga, USA, November 2013.

[9] G. D. F. Morales, A. Gionis, and C. Lucchese, "From chatter to headlines: harnessing the real-time web for personalized news recommendation," in *Proceedings of the 5th ACM International Conference on Web Search and Data Mining (WSDM ’12)*, Seattle, Wash, USA, February 2012.

[10] F. Abel, Q. Gao, G.-J. Houben, and K. Tao, "Twitter-based user modeling for news recommendations," in *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI ’13)*, pp. 2962–2966, AAAI Press, Beijing, China, August 2013.

[11] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. Magnenat-Thalmann, "Who, where, when and what: discover spatio-temporal topics for twitter users," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD ’13)*, pp. 605–613, Chicago, Ill, USA, August 2013.

[12] M. Lichman and P. Smyth, "Modeling human location data with mixtures of kernel densities," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD ’14)*, pp. 35–44, ACM, New York, NY, USA, August 2014.

[13] K. Lee, R. K. Ganti, M. Srivastava, and L. Liu, "When twitter meets foursquare: tweet location prediction using foursquare," in *Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MOBIQUITOUS ’14)*, pp. 198–207, ICST, London, UK, December 2014.

[14] B. Han, P. Cook, and T. Baldwin, "Textbased twitter user geolocation prediction," *Journal of Artificial Intelligence Research*, vol. 49, pp. 451–500, 2014.

[15] W. Mathew, R. Raposo, and B. Martins, "Predicting future locations with hidden Markov models," in *Proceedings of the ACM Conference on Ubiquitous Computing (UbiComp ’12)*, Pittsburgh, Pa, USA, September 2012.

[16] J. Bao, Y. Zheng, and M. F. Mokbel, "Location-based and preference-aware recommendation using sparse geo-social networking data," in *Proceedings of the 20th International Conference on Advances in Geographic Information Systems (SIGSPATIAL ’12)*, pp. 199–208, ACM, Redondo Beach, Calif, USA, November 2012.

[17] Q. Yuan, G. Cong, and A. Sun, "Graph-based point-of-interest recommendation with geographical and temporal influences," in *Proceedings of the 23rd ACM International Conference on Information and Knowledge Management (CIKM ’14)*, pp. 659–668, ACM, Shanghai, China, November 2014.

[18] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. Magnenat-Thalmann, "Time-aware point-of-interest recommendation," in *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR ’13)*, pp. 363–372, 2013.

[19] H. Gao, J. Tang, X. Hu, and H. Liu, "Exploring temporal effects for location recommendation on location-based social networks," in *Proceedings of the 7th ACM Conference on Recommender Systems (RecSys ’13)*, pp. 93–100, Hong Kong, October 2013.

[20] J. W. Pennebaker, C. K. Chung, M. Ireland, A. Gonzales, and R. J. Booth, *The Development and Psychometric Properties of LIWC2007*, LIWC, Net, Austin, Tex, USA, 2007.

[21] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: LIWC and computerized text analysis methods," *Journal of Language and Social Psychology*, vol. 29, no. 1, pp. 24–54, 2010.

[22] A. D. I. Kramer and C. K. Chung, "Dimensions of self-expression in facebook status updates," in *Proceedings of the 5th
[23] J. Chen, G. Hsieh, J. Mahmud, and J. Nichols, “Understanding individuals’ personal values from social media word use,” in Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW ’14), pp. 405–414, ACM, San Francisco, Calif, USA, February 2014.

[24] S. A. Golder and M. W. Macy, “Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures,” Science, vol. 333, no. 6051, pp. 1878–1881, 2011.

[25] S. Argamon, M. Konnel, J. W. Pennebaker, and J. Schier, “Mining the blogosphere: age, gender and the varieties of self-expression,” First Monday, vol. 12, no. 9, 2007.

[26] E. Gilbert, “Phrases that signal workplace hierarchy,” in Proceedings of the Conference on Computer Supported Cooperative Work (CSCW ’12), pp. 1037–1046, Seattle, Wash, USA, February 2012.

[27] J. Mahmud, M. X. Zhou, N. Megiddo, J. Nichols, and C. Drews, “Recommending targeted strangers from whom to solicit information on social media,” in Proceedings of the International Conference on Intelligent User Interfaces (IUI ’13), pp. 37–47, ACM, March 2013.

[28] K. Lee, J. Mahmud, J. Chen, M. X. Zhou, and J. Nichols, “Who will retweet this?: automatically identifying and engaging strangers on twitter to spread information,” in Proceedings of the 19th International Conference on Intelligent User Interfaces (IUI ’14), pp. 247–256, ACM, Haifa, Israel, February 2014.

[29] H. Badenes, M. N. Bengualid, J. Chen et al., “System U: automatically deriving personality traits from social media for people recommendation,” in Proceedings of the 8th ACM Conference on Recommender Systems (RecSys ’14), pp. 373–374, Foster City, Calif, USA, October 2014.

[30] H. A. Schwartz, J. C. Eichstaedt, M. L. Kern et al., “Personality, gender, and age in the language of social media: the open-vocabulary approach,” PLoS ONE, vol. 8, no. 9, Article ID e73791, 2013.