2014 IEEE International Conference on Data Mining Workshop

Why Checkins: Exploring User Motivation on Location Based Social Networks

Fengjiao Wang, Guan Wang, Philip S. Yu
Department of Computer Science
University of Illinois at Chicago
Chicago, Illinois, 60607, USA
Email: {fwang27, gwang26, psyu}@uic.edu

***Abstract*—Checkins, the niche service provided by location based social networks (LBSN), bridge users' online activities and offline social lives in a seamless way. Therefore, knowledge discovery on checkin data has become an important research direction [1], [2], [3], [4]. However, a fundamental and interesting question about checkins remains unanswered yet. What are people's motivations behind those checkins? We give the first attempt to answer this question. Motivation studies first appear in social psychology in a less quantitative way. For example, the goal-directed behavior (MGB) model [5] uncovers the association between behaviors and motivations. Following a similar rationale, we design a computational model for the mining of user checkin motivations from large scale real world data. We assume that the checkin motivation has two types: social motivation and individual motivation. Social motivation is the type of checkin incentive that stimulates interactions or influences among friends. Individual motivation is another type of checkin incentive that aims to explore and share attractive places. Following the structure of the MGB model, we construct user checkin motivation prediction model (UCMP) and then formalize the motivation prediction problem as an optimization problem. The idea is minimizing the difference between the estimated behavior and the true behavior to get the predicted motivations. The experiment on this GOWALLA dataset shows not only prediction results, but also very interesting phenomenons about social users and social locations.**

I. Introduction

The location-based social networks (LBSNs) such as Gowalla, Foursquare and Brightkite offer a unique behavior, *checkin*, to enable the location based social experience. By exploring LBSN, users can record where they are, interact with their friends or share and comment friends' checkin information. Therefore, LBSN bridges users' offline activities and online interactions and becomes a flourishing social platform. Recently, more and more researchers are paying attentions to LBSN data to discover knowledge related to people's online and offline lives. For example, interesting studies have been done to discover human mobility patterns and predict future check-ins [6], [7], [8], uncovering travel patterns [9], [10] to facilitate traveling guidance, and detecting socio-spatial properties of location-based social networks to infer friendship [2]. Despite the series of findings, a fundamental question seems to be neglected and unanswered so far, which is the motivation behind every checkin. Previous work largely focused on using checkins to discover new knowledge, or predict possible checkins. We believe that if we can discover the motivation behind each checkin, there will be great benefits to the existing tasks. For example, checkin prediction can be more accurate. If we know the reasons why a user want to check in at different places, given a new location we can better predict whether the user will check in at this location or not. Another example, with better understanding of checkin motivation, we can improve the advertisement effect. Instead of targeting the advertisement messages to all of the people, the shop owners would like to target to the users who probably will visit their shops after seeing the advertisement messages. If users usually behave very socially, once knowing their friends check in some new places, they may be very willing to try this new place by themselves or hangout with their friends. Targeting the advertisement messages to this type of users will get more benefits than others. Therefore, we give the first shot to provide a comprehensive study of the user's check-in motivations in this paper.

We develop a checkin motivation mining model based on the concepts from the social psychology field where human behavior modeling has been an established field for a long time [5], [11], [12]. In social psychology, behavior study is mainly theoretical with small scale experiments, due to the limitation of available data. Usually, the studies can only be conducted by making surveys, which can't fully reflect the properties in a large data set. The availability of large amounts of checkin information in location-based social networks provides user behavior data in a large scale. It is worthy to combine the established concepts in social psychology with computational models for knowledge discovery on a large scale dataset. In our model, we are inspired by the MGB model in behavior analysis in social psychology. Specifically, the MGB model concludes that attitudes towards the act, subjective norms and perceived behavior controls are predictors of desires (motivation). Desires (motivation) are proximal causes of intentions. In addition, intentions measure a person's relative strength of willingness to perform a behavior. (Figure 2). Mapping to the checkin scenario, attitudes correspond to the user's favor toward checking in place, subjective norms correspond to the influence from friends about checking in place, and perceived behavior controls correspond to the limitation that the user perceived when performing checkin action. Therefore, the three predictors in MGB model can be mapped to the features in our learning model. We thus discover the mappings between these social science concepts and the features in LBSN for checkin motivation discovery.

As we assumed, the User Checkin Motivation Prediction Model (UCMP) contains three features, user's attitude, subjective norm, and perceived behavior control. In addition, checkin behavior is known given checkin information. In order to predict checkin motivation (Desires), we formalize

978-1-4799-4274-9/14 $31.00 © 2014 IEEE
DOI 10.1109/ICDMW.2014.175

the prediction problem as an optimization problem. It acquires the optimized motivation values by minimizing the difference between estimated behavior and actual behavior. In order to get a better prediction result, we introduce two constraints. They minimize the difference between the two motivations and observed information. The optimization problem is solved using gradient descent method.

Our Contributions With our analysis on co-checkin information on LBSN and experiment results, some interesting phenomena about users' checkin motivations are observed. At the same time, we can perform promising applications with checkin motivation information.

1) Co-checkin Information: co-checkin information can be used to indicate social users and social locations.
2) Checkin Motivation Study: to the best of our knowledge, this is the first time we use location-based social networks to study user checkin motivations in a large scale dataset. This study can benefit social psychology studies to fill in the blanks of large scale studies in social psychology field. It can also benefit many practical applications.
3) Motivation Model: we use the social psychology model MGB to capture the formation process of behavior, and keep a positive relationship in the construction to predict user checkin motivations using UCMP model.

II. GOWALLA DATA

A. Data

The dataset used in this paper comes from previous LBSN research [1]. It contains friendship network with 196,591 users and 950,327 friend links. It also has a collection of 6,442,890 check-ins from Feb. 2009 to Oct. 2010. For the location, the dataset only provides latitude and longitude. In order to fulfill evaluation purpose, we collect another GOWALLA dataset containing geometric information of 30367 locations.

B. Co-checkin Analysis

We define co-checkin as when user and his friend check in at the same location within a short period (we use one week as the time range). One possible reason co-checkin phenomenon exists is that user influences checkin decision of his friends by checking in at the same place earlier. We define checkins influenced by friends are socially motivated checkins. Therefore, co-checkin phenomenon is an important factor to indicate checkin motivations. In the following part, we show several interesting observations of co-checkin information (Figure 1(a) to Figure 1(f)). Specifically, we are interested in 2 questions:

1) Is co-checkin phenomenon prevalence among users or not? The answer to this question shows to what extent that the co-checkin information can help uncover the checkin motivations. (Figure 1(a)).
2) Do users have the same or different preferences towards co-checkin? The answer to the two problems guides us in building motivation prediction model. (Figure 1(d) to Figure 1(f)).

Figure 1(a) answers the first question. It shows the histogram of co-checkin percentage (i.e., number of users whose co-checkin percentage falls into different categories). For most users, whose co-checkin percentage is smaller than 10%. However, there are a certain number of users whose co-checkin percentage falls between 10% and 80%. Co-checkin phenomenon important for motivation prediction, but it is not sufficient. Therefore, we utilize co-checkin information, friendship network and checkin information in the proposed model.

In Figure 1(b) to Figure 1(f), we show more observations on Gowalla data. Number of checkins is illustrated on X-axis, user proportion or number of users is illustrated on Y-axis. To clearly deliver the message, two scales are applied in horizontal axis of these figures, divided by a red vertical bar. Figure 1(d) shows the histogram of check-ins, where blue bar shows number of users fall into different categories, red bar shows number of users who have zero co-checkin fall into different categories. We can conclude that the less checkin the user has the more likely the user will have zero co-checkin. Interestingly, some users do not have co-checkin even if they have 200 or 300 total checkins. It may indicate that users may have different preferences for co-checkins.

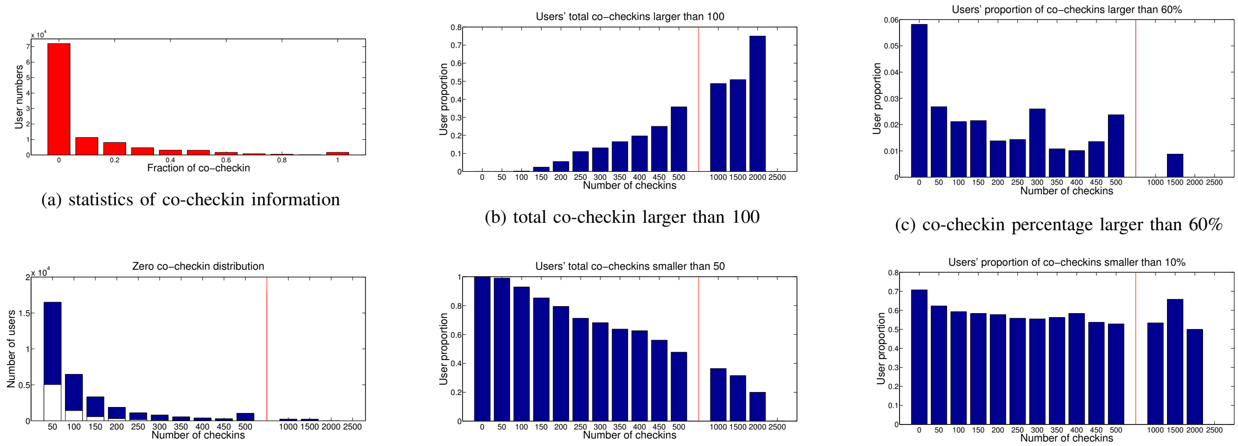
In Figure 1(b) and Figure 1(e), user proportion is illustrated in Y-axis. Specifically, the height of the bar in Figure 1(b) equals to the proportion of users whose total number of co-checkins larger than 100. The height of bar in Figure 1(e) equals to the proportion of users whose total number of co-checkins smaller than 50. Two figures indicate that people have more checkins, they are more likely to have a large number of co-checkins and less likely to have a small number of co-checkins. Blue bar in Figure 1(c) shows the proportion of users whose percentage of co-checkins larger than 60%. Blue bar in Figure 1(f) shows the proportion of users whose percentage of co-checkins smaller than 10%. When the users have more checkins, user proportion in Figure 1(c) and Figure 1(f) is not increase or decrease monotonically. It implies that users have different preferences toward co-checkin. User may have zero co-checkins even if he has a large number of checkins. Based on the above observations, we can conclude that users can be categorized into different types. Some users may prefer to checkin for social purpose, and others may not. From the above analysis, we can confirm that co-checkin information is useful for checkin motivation study. However, the relation between co-checkin and checkin motivation is complicated, i.e., the relation can not be summarized by a single rule.

III. USER MOTIVATION MODEL AND ALGORITHM

In the previous section, we define the motivation prediction problem and introduce how to map the social psychology concepts of MGB model to the features of user checkin scenario. In this section, we make an observation of the checkin behavior which prove that the features mapped from predictors of MGB model are necessary for solving motivation prediction problem. Then, we introduce the details of the User Checkin Motivation Model (UCMP) and the algorithm.

A. Problem

Before we present a formal definition of the problem, we first give several necessary notations. A location-based social network can be represented as $G = (U, F, L, C)$, where U is



(a) statistics of co-checkin information

(b) total co-checkin larger than 100

(c) co-checkin percentage larger than 60%

(d) zero co-checkin distribution

(e) total co-checkin smaller than 50

(f) co-checkin percentage smaller than 10%

Fig. 1: Co-checkin analysis

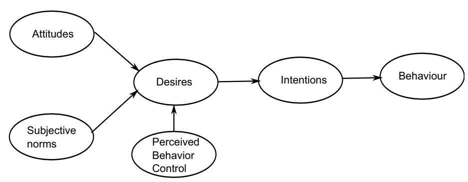


Fig. 2: Simplified model of goal-directed behavior

the set of $|U| = M$ users, $F \subset U \times U$ is the set of links representing friend relationship between users, L is the set of $|L| = N$ locations, and $C \subset U \times L \times T \times M$ is the set of check-in tuples. Each check-in tuple c_i is represented as $c_i = (u_j, l_k, t, m_i)$, which means user u_j check in at location l_k at time t with motivation m_i. The motivation prediction task is to use the information in the location-based social network G to predict each check-in c_i's motivation m_i.

B. Model of goal-directed behavior and its interpretation

Behavior modeling problem has been studied for a long time in social psychology field. MGB model is a model of purposive behavior which introduce desires (motivation) as a new variable, because desires are shown to play an important role in decision-making. The MGB model suggest that: desires are the proximal causes of intentions; desires also transform the motivational content to act embedded in behavioral attitudes (BA), subjective norms (SN) and perceived behavior control (PBC) (Figure 2). It provides us a structure that links different social psychology concepts, and can be utilized to predict motivation. However, the model can't be directly utilized in checkin scenario because it was used to handle very small dataset (e.g. survey data). Therefore, we want to design a computational model which contains such a structure and also applicable to large scale dataset. Before we go to details of the computational model, we need to map the social psychology concepts to checkin scenario firstly.

Definition 3.1 (Behavioral Attitudes (BA)): An individual's positive or negative evaluation of self-performance of the particular behavior is called behavioral attitude.

In one word, behavior attitude means an individual's attitude toward the particular behavior. In the checkin scenario, it means that whether the user like to check in the place or not. We assume that there are two motivations (social motivation and individual motivation) with regards to checkin behavior. In the MGB model, motivation transform the motivational content embedded in BA, SN, PBS to act. Since there are two motivations, we extract two set of motivational contents from all predictors (BA, SN, PBC). For the social motivation, behavior attitude (denoted by $\alpha^0_{u_j}$) means that whether the user (u_j) like checking in places for the purpose of interacting with his friends. For the individual motivation, behavior attitude (denoted by $\alpha^1_{u_j}$) means that whether the user (u_j) like checking in places attracted by the location.

Definition 3.2 (Subjective Norm (SN)): An individual's perception of social normative pressures, or relevant others' beliefs that he or she should or should not perform such behavior is called subjective norm.

Subjective Norm means the pressure that the individual received from his/her friends. In our cases, subjective norm (denoted by $\gamma^{l_k}_{u_j}$) means that whether user (u_j) feel external pressures or receive influences from his friends to check in the place (l_k).

Definition 3.3 (Perceived Behavioral control (PBC)): An individual's perceived ease or difficulty of performing the particular behavior is called perceived behavior control.

Perceived behavior control is the feeling of individual about the difficulty of performing the particular behavior. There can be many different difficulties, for example difficulty of employing a particular app, networking issue, cost and restriction of location. For the checkin information, the restriction of location is the one we can measure. Perceived behavior control in our scenario means that whether the place is suitable for user check in with different purposes. It also contains two factors: for social motivation we use perceived behavior control ($\beta^0_{l_k}$) to indicate whether the place (l_k) is suitable for social purpose checkin; for individual motivation, we use perceived behavior control ($\beta^1_{l_k}$) to indicate whether the place (l_k) is suitable for individual purpose checkin.

Definition 3.4 (Desires): An individual's reason for performing the particular action is called Desires. Desires (de-

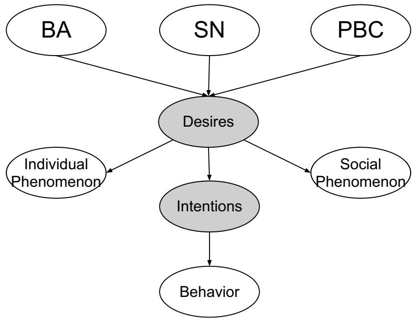


Fig. 3: Prototype of UCMP model

noted by m_i) are modeled as two types of motivations (social motivation and location motivation) that drive the checkin behavior.

Definition 3.5 (Intentions): An indication of an individual's readiness to perform a given behavior is called intention. In our cases, it (denoted by n_i) measures how likely that desires stimulate the behavior of check-in. In other words, intention means the possibility of user performing the checkin action stimulated by desires (motivations).

Definition 3.6 (Behavior): An individual's observable response in a given situation with respect to a given target is called behavior. In our cases, behavior (denoted by b_i) is user performed checkin action.

Definition 3.7 (Individual Phenomenon): It tests whether the checkin is unique checkin, which means the place is not checked by his friends and himself earlier.

Definition 3.8 (Social Phenomenon): It tests whether the checkin is co-checkin, which means his friends checked into the same place within one week.

Besides mapping the variables represented in the simplified MGB model (Figure 2), we introduce two phenomenons to better predict the check-in motivation. Specifically, it contains whether the checkin is unique checkin and whether the checkin is co-checkin. The Figure 3 explains all the variables utilized to capture the internal relations and predict user motivations.

C. User Motivation Model

The motivations we want to predict correspond to the desires in MGB model. In order to predict motivations, we need to utilize all the predictors (behavior attitude, subjective norm and perceived behavior control) which determine the motivations. In addition to that, we need the observations (behavior, individual phenomenon and social phenomenon) which are determined by motivations (Figure 3) to form the optimization problem. The objective of the optimization problem is to minimize the difference between the estimated behavior and the actual behavior. By solving the optimization problem we can get the predicted motivations.

Firstly, we give the notations of the variables explained in last section in Table 1.

Secondly, we explain how to calculate the features. Before we describe the details of explaining the variables, we first introduce several useful notations of this part.

Variable	Factor	Notation
Behavior Attitude	$UT1$ $UT2$	$\alpha^0_{u_j}$ $\alpha^1_{u_j}$
Subjective Norm		$\gamma^{l_k}_{u_j}$
Perceived Behavior Control	$LT1$ $LT2$	$\beta^0_{l_k}$ $\beta^1_{l_k}$
Desires	motivation	m_i
Intention		n_i
Behavior		b_i
Individual Phenomenon	unique check-in	p^0_i
Social Phenomenon	co-checkin	p^1_i

TABLE I: Notations of Prediction variables

Notation	Definition
C_{u_j}	user u_j's total checkins
C_{l_k}	location l_k's total checkins
c^0_i	check-in c_i is co-checkin
c^1_i	check-in c_i is unique check-in
$C^0_{u_j}$	the set contains user u_j's total co-checkin
$C^1_{u_j}$	the set contains user u_j's total unique check-in
$C^0_{l_k}$	the set contains location l_k's total co-checkin
$C^1_{l_k}$	the set contains location l_k's total unique check-in
F_{u_j}	the set contains all the friends of user u_j
$F^{l_k}_{u_j}$	the set user u_j's friends who have checked in location l_k

The factors belong to the variable on behavior attitude explains the user's attitude towards check-in behavior. We push our study further, instead of studying the user's attitudes toward check-in behavior, we study the user's attitudes towards check-in for social motivation or location motivation. We claim that the attitude of checkins drived by the two types of motivations are totally different. For example, some people may just use the LBSN as a location recorder. They post check-in information because they want to record where they have been to and mark the interesting places. They don't use it for social interactions or maybe they dislike it. In this case, the user's attitudes toward check-in for location motivation is highly positive, but the attitudes toward check-in for social motivation is negative. Distinguish the two attitudes can help us better predict checkin motivations.

The definition for the first factor $\alpha^0_{u_j}$ is user u_j's attitudes toward check-in for social motivation. Specifically, the higher percentage of user's co-checkin the attitude towards social motivation check-in is more positive. $\alpha^0_{u_j} = \frac{\#C^0_{u_j}}{\#C_{u_j}}$, where $C^0_{u_j}$ denotes user u_j's total co-checkin and C_{u_j} denotes user u_j's total checkins. Accordingly, $\alpha^1_{u_j}$ is defined as, $\alpha^1_{u_j} = \frac{\#C^1_{u_j}}{\#C_{u_j}}$, where $C^1_{u_j}$ denotes user u_j's total unique checkins and C_{u_j} denotes user u_j's total checkins.

The second factor is the user's friends influence to his check-in at the location. We use a simple assumption here, which is each time the user's friend post a check-in information, the user can see this information on his own page. The assumption is true for LBSNs such as Gowalla, Foursquare. More of his friends check-in at the same place, the larger influence the user can receive from his friends. We just consider the positive influence. The friend influence is defined as, $\gamma^{l_k}_{u_j} = \frac{\#F^{l_k}_{u_j}}{\#F_{u_j}}$ where $F^{l_k}_{u_j}$ denotes user u_j's total number of friends who have checked in at location l_k and F_{u_j} denotes the total number of friends of user u_j.

The third factor is perceived behavior control which means how user's behavior is controlled or limited. With respect to the check-in behavior, the checkin behavior is probably restricted by location. In other words, how a location can restrict the user's check-in motivations is considered in this factor. It corresponds to two factors: first, whether the location is suitable for social motivation; second, whether the location is suitable for location motivation. For the shorthand, we denote the two factors of feature perceived behavior control as first type ($LT1$) and second type ($LT2$). The likelihood of $LT1$ and $LT2$ are as follows, $\beta^0_{l_k} = \frac{\#C^0_{l_k}}{\#C_{l_k}}$ where, $C^0_{l_k}$ denotes location l_k's total number of co-checkin and C_{l_k} denotes location l_k's total number of checkins. $\beta^1_{l_k} = \frac{\#C^1_{l_k}}{\#C_{l_k}}$, where, $C^1_{l_k}$ denotes location l_k's total number of unique checkin and C_{l_k} denotes location l_k's total number of checkins.

For ease of computation, we treat behavior attitude as vector α_{u_j} which contains $\alpha^0_{u_j}$ and $\alpha^1_{u_j}$. And also for the variable on subjective norm, denoted as β_{l_k}.

In order to better explain behavior's formation process, we introduce three functions (Figure 3). Desires is the function of behavioral attitude, subjective norm and perceived behavioral control. Intention is the function of desires. The behavior is the function of intention.

For the function of desires, previous meta-analysis research ([11]) for the goal-directed model shows that attitude have strongest prediction ability for desire, subjective norm have less prediction ability than attitude, and perceived behavior control is a necessary factor in determining desires. Overall, the contributions of the three variables to predict desires are different. Therefore, we use different weight for different variables. The desires function f is modeled as linear combination of all the predictive factors.

$$m_i = w_0 + w_1 * \alpha_{u_j} + w_2 * \gamma^{l_k}_{u_j} + w_3 * \beta_{l_k} \tag{1}$$

(same formation with reference [12]). In function f, α_{u_j} is user u_j's attitude towards check-in behavior and $\gamma^{l_k}_{u_j}$ is user u_j's perceived influence from his friends with respect to location l_k. All arguments' range of function m_i is $[0, 1]$. The same users' different check-ins will share the same attitude and subjective norm. β_{l_k} is perceived behavior control, in our scenario which means the location l_k's limitations to the check-in motivation, so the different check-ins in the same location will share the same β_{l_k} value. According to the previous definition, desire m_i is a vector, where the first element m^0_i is used to predict social motivation and the second m^1_i is used to predict location motivation.

For the intention function, different types of desires have different influence on intention. And also, the stronger the desires is, the stronger the intention will be. With the two knowledge of two types of motivations, we design the intention function as a logistic function.

$$n_i = \frac{2}{1 + e^{-(m^0_i + m^1_i)/2}} - 1 \tag{2}$$

The value of behavior is equal to intention n_i. The values of estimated individual phenomenon and estimated social phenomenon equal to m^0_i and m^1_i.

In order to predict the motivation, we first estimate motivations using Equation 1, then get the motivation values by iteratively changing the estimated motivation value and minimizing the difference between estimated behavior and actual behavior. At the same time, the optimization must satisfy the two constraints. Since the individual phenomenon and social phenomenon can be indicated by two motivations, we use the difference between actual phenomenons and predicted phenomenons as constraints. Specifically, the constraints reflect that the difference between the observed phenomenon and the estimated phenomenons should not be too much. The objective function F is defined as follows,

$$\mathrm{F} = \min(\|p^0 - m^0\|_2 + \|p^1 - m^1\|_2 + \|b - n\|_2) \tag{3}$$

D. Algorithm

We utilize the gradient descent method to minimize the objective function.

The algorithm is as follow:

Data: behavior attitude factor α, subjective norm factor γ, perceived behavior control factor β, phenomenon p, behavior factor b and learning rate δ
Result: checkin motivation m^0_i, m^1_i
initialization $w_0, w_1, w_2, w_3, \delta, i = 0$;
Compute motivation variables m^0_i, m^1_i;
repeat
  Compute gradient $\nabla_{w_0}, \nabla_{w_1}, \nabla_{w_2}, \nabla_{w_3}$;
  update w_0, w_1, w_2, w_3 value;
  $w_0 = w_0 - \delta * \nabla_{w_0}$;
  $w_1 = w_1 - \delta * \nabla_{w_1}$;
  $w_2 = w_2 - \delta * \nabla_{w_2}$;
  $w_3 = w_3 - \delta * \nabla_{w_3}$;
  Compute motivation variables' values m^0_i, m^1_i;
  Compute objective value F;
  i++;
until *F not change or i==1000*;

Algorithm 1: Gradient Descent Algorithm

IV. EXPERIMENT

We demonstrate our results using motivation prediction task on real world dataset. In order to show the practical potential of motivation study, we tackle future checkin prediction task on another experiment.

A. Motivation Prediction

In this section, we first present two case studies to demonstrate the effectiveness of the proposed approach. Next, we present the aggregate experiment results. Finally, we introduce the interesting observations obtained from motivation prediction results. In the previous sections, we assume there are two motivations: social motivation and individual motivation. If one user have many social motivation checkins, we will name the user as social user. The same definition techinique applies to location. The location will be named as social location if many social motivation checkins take place in this location.

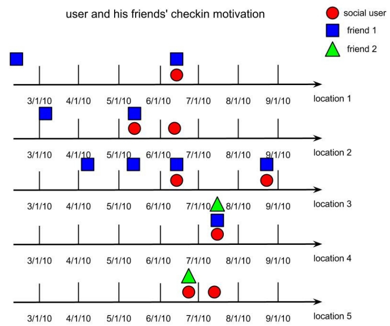


Fig. 4: User and his friends' checkin motivation

1) Case Study: Now, we introduce two case studies to demonstrate the effectiveness of the proposed approach. The two case studies are respectively targeting at social users and social locations. Figure 4 displays an example from the experiment. It represents a portion of the checkins of a social user and his friends. The user has 225 checkins, among the total 225 checkins 196 of them are socially motivated. The percentage of social motivation checkins of this user is 87%. Red circles in Figure 4 denote the user's checkins. Blue squares correspond to friend 1's checkins, and green triangles are associated with checkins of friend 2. Each horizontal ordinate denote one unique location, with mark indicate the time stamp. Each shape located above one horizontal ordinate and between two marks means the checkin performed at that location and between those two time stamps. For example, if friend 1 checked in at location 1 on $2/15/2010$, there will be one blue square above the first horizontal ordinate before the time stamp $3/1/2010$. In Figure 4, all checkins of this social user are predicted to social motivation checkin. These checkins can be categorized into two types: the user performed checkin in the same location at the same day with his friend; the user performed checkin again after he and his friend checked in the same location at the same day. For example, the user performs checkins at location 1 on $6/13/2010$; location 2 on $5/11/2010$; and location 3 on $6/13/2010$ and $8/24/2010$ belong to the type 1. The user performs checkins at location 2 on $6/10/2010$ and location 5 on $7/14/2010$ belong to type 2. For the first type of checkins, the user and his friend check in at the same place on the same day which indicates that the user and his friend may visit this place together. In addition, the user and his friend may visit some other places together. This type of checkins facilitate social connection, so they are socially motivated. For the second type of checkins, the user check in the place soon after he and his friend check in this place together. It indicates that either the user is influenced by his friend or he wants to connect with his friend by checking in this place. Whether checking in places is for the purpose of social contact or is the result of social influence, the checkin is socially motivated. Social user is the user who has many checkins which are socially motivated. Since 87% percent of the user's checkins are predicted as social motivation checkins, we categorize the user as social user. The reason why the user is social users is that he checks in different places together with his different friends many times. If he and his friend checkin together one time, it is an occasional case, which can't conclude that the user is always socially motivated. However, if the user checkin together with his different friends many times, the user probably is a social user. The first case study demonstrate that social motivation checkin and social user are predicted correctly.

Fig. 5: Case Study - Location

The second case study shows a small portion of locations. By utilizing the location information contained in the second GOWALLA dataset, we draw several locations in Google Map in Figure 5. Each pin represents a location, with red pin denoted social location and blue pin denoted non-social location. Locations A,B,C,D,E are five Apple Stores; location F is Starbucks Coffee; location G is Target; and location H is John Adams Building. Since starbucks Coffee is the place where people may hangout together, it is a social location. Target is a department store where people buy what they need, so it is not a social location. John Adams Building is an office building, which is not a social location. Besides clearly differentiating the social places for different types of locations, the algorithm is able to distinguish the same type of places at different locations. The three Apple Stores located in Manhattan have the higher percentage of social motivation checkins than the other two. The Apple store in Paramus has no social motivation checkins. It is likely due to that people live in a suburb like Paramus are couples with kids, while a lot of people live in Soho or NY city are young professionals who are still singles or married without kids, and hence are more likely to hangout at Apple stores. Apple store may not be the place that people live in Paramus would like to hangout, as they may prefer other entertainment. For the three Apple stores located in Manhattan, the percentage are also different. Apple store in Soho has the highest percentage. The second case study demonstrates that the algorithm is able to discover social locations.

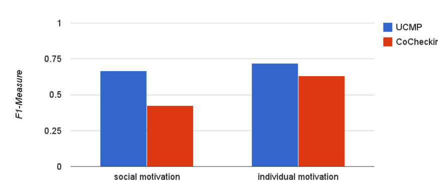


Fig. 6: Experiment Result

2) Motivation Prediction Task: Based on the UCMP model (Section 3), we formalize the motivation prediction problem as an optimization problem. The algorithm predicts the motivation of each checkin on GOWALLA dataset. 11.4% of checkins are predicted as social motivated, while the other

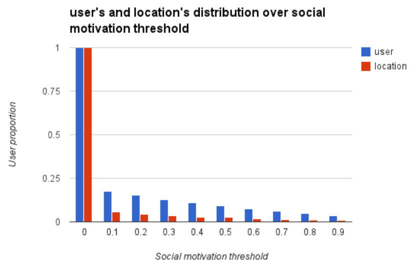


Fig. 7: user's and location's distribution over percent of social motivation

88.6% of checkins are predicted as individually motivated. We randomly pick 236 checkins, and invite the human to label the motivation of each checkin. Figure 6 shows the comparison result of the proposed method and the baseline Co-Checkin. Baseline Co-Checkin considers only the co-checkins as socially motivated and others are individually motivated. For both the social and individual motivations, the proposed algorithm UCMP performs better than the baseline. Baseline Co-Checkin considers only the checkins which are co-checkins. However, the previous case study shows that the checkins which are the consequence of social influenc also should be considered as socially motivated checkin. Therefore, the proposed method performs better than the baseline.

3) Observation: Each checkin is performed by one user at one location. Knowing the motivations of all checkins can facilitate the study of user's general checkin motivation and location's general checkin motivation. It can aid many practical applications, such as checkin prediction, advertisement recommendation and other applications. Figure 7 shows the faction of users/locations fall into different categories according to social motivation threshold. User's social motivation value is the percentage of user's social motivation checkins. For example, when the social motivation threshold equals to 0.3, Figure 7 calculates fraction of users/locations whose social motivation value larger or equal to 0.3. 18% of users whose social motivation values are larger than 10%, and 6% of locations whose social motivation value are larger than 10%. The fraction of users is higher than the fraction of locations with social motivation value larger or equal to 10%. It is likely due to the fact that the average number of users' checkins is much smaller than the average number of locations' checkins, so users are easier to have large percentage of social motivation checkins. In addition, The locations' checkin motivation can be very diversified, so it is hard to find some locations whose percent of social motivation checkins are very high. There are users and locations have high percentage of social motivation checkins, but also users and locations have low percentage of social motivation checkins. It means users and locations have different types with respect to social motivation checkins.

The Figure 7 reflects that users have different types. However, it is not clear that what kind of users are social users and what kind of locations are social locations. One characteristic of social users is they have more checkins in the same places. The average number of checkins per location of social users is 1.97, while for all the other users is 1.50. It means that for social users they at least check in the same place twice in average. The same thing happens for social locations. The average number of checkins per user of social locations is 2.36, while for all the other locations is 1.67. More interestingly, the motivation data shows that social users are more likely to checkin social locations. We randomly pick a social user, among all the locations the user checked in the average percentage of social motivation checkins is 26%. It is a high number compared with 11%, which is the percentage of social motivation checkins among all users. We also randomly pick a social location, almost all of the users checked in at this location have higher percentage of social motivation checkins.

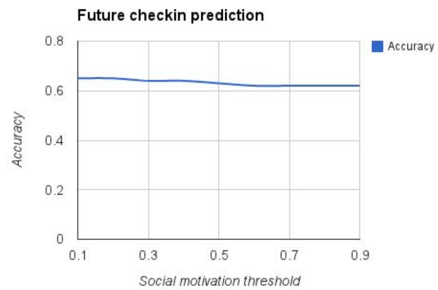


Fig. 8: prediction accuracy of user's future checkin

B. Future Checkin Prediction

In the case study section's examples, when a user hangs out with his friends in some places, he checks in this location later by himself. The motivation of the user's later checkin is social motivation, because the user may also wants to co-checkin with his friends or he is influenced by his friends. He may also hang out with his other friends in this place or influence his other friends to visit this place. It gives us a hint that the user's future checkin may be related to user's previous checkin motivations.

1) Whether the user checkin in the future or not: We use GOWALLA dataset to perform the prediction task. There are 20 month checkin information in this dataset. We use previous 19 month checkins to predict the checkins of the last month. There are certain amount of new checkins (checkins performed by users who did not check in before) each month. It's hard to predict users who did not check in before, since we do not have any information of the users. We simply delete the last month's checkins. The prediction method utilizes three kind of information: user's previous checkin, friends' previous checkin, and user's social motivation value. The user's social motivation value can help decide whether the user will check in at this place given the user's friends already checked at this place. If the user's friend checked in at this place before and also the user is very social (the user like to interact with their friends by checking in places), it is likely that user will check in at this place later. Therefore, we use social motivation value as threshold to predict the future checkin. For example, if user's social motivation value is larger than threshold 0.3 and also user's friend check in at this place before, we predict the user will check in this place in the future. The accuracy is the combined accuracy of checkin and non-checkin cases. The overall accuracy is 65%.

2) Whether future checkin is socially motivated or not: We only utilize the previous information to predict the motivations of future checkins. The previous information contain user's previous checkins, friends' previous checkins and user's social motivation value. Figure 9 shows the prediction results by using the results of motivation prediction part (section 3)

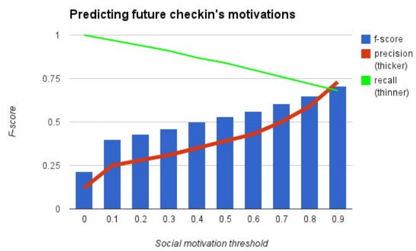


Fig. 9: precision, recall, F-score of future cocheckin prediction

as benchmark. The thinner green curve represents the recall; the thinker red curve denotes the precision. The F-score is represented as blue bar. The figure indicate that when the social motivation threshold is 0.9, the experiment gets the best prediction result. It means that when predicting future checkin motivations, using the large social motivation value will get the best result.

V. Related Work

A. User Intention Modeling

The relation between attitudes and behavior is a central theme in modern social psychology. The researchers distinguish three generations of attitude-behavior relation. Among many models which explained how the attitude leads to behavior, Fish-bein and Ajzen's theory of reasoned action ([13]) attracted most attentions. It stimulate many later studies not only on the meta-analysis ([11]) but also the practical studies. Two advantage of the theory is its parsimony and applicable to many fields. Though it has two advantages, the model can't be applied to purposive behaviors. The model of goal-directed behavior ([5]) studied the purposive behavior by adding more prediction variance and introducing a new construct which explains the existing predictors. It is suitable for modeling goal-directed behaviors. As checkin behavior is pervasive in location-based social networks, it can be viewed as a goal-directed behavior. To explore the place or interacting with their friends, the user choose to check in at this place. Exploring the motivation behind check-in behavior is an interesting yet untouched field.

B. Check-in behavior study

As location-based social networks ([14], [15], [16], [17], [18]) provide user check-in and online communication, more and more users are attracted to location-based social networks. Users post where the places they are on the location-based social networks and add photos, and tips to communicate with friends online or win rewards or get discounts. Location-based social networks provide unique data sets which contain both offline activities and online interactions. It is for this reason, researchers are able to study why users want to use location-based social networks ([19], [20]), and why users are happy to share their check-in information with friends.

VI. Conclusion

This paper gives the first try to tackle the motivation prediction problem in LBSNs. It is not only a significant problem for user behavior analysis, but also a potential benefit for some other applications. Inspired by the previous social psychology studies, we construct the behavior formation model and formulate the motivation prediction problem as an optimization problem. Possible future extensions include two directions. On one hand, deeply understanding the diversified motivations on a large scale data set will benefit the user behavior studies and also the social psychology community. On the other hand, extending the existing framework to improve the results of potential applications such as targeting advertising and recommendation also worth trying.

References

[1] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: Friendship and mobility: User movement in location-based social networks," *KDD*, 2011.

[2] J. Chang and E. Sun, "Location3: How users share and respond to location-based data on social," *ICWSM*, 2011.

[3] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil, "An empirical study of geographic user activity patterns in foursquare," *ICWSM*, 2011.

[4] S. Scellato, A. Noulas, R. Lambiotte, and C. Mascolo, "Socio-spatial properties of online location-based social networks," *ICWSM*, 2011.

[5] M. Perugini and R. Bagozzi, "The role of desires and anticipated emotions in goal-directed behaviours: Broadening and deepening the theory of planned behavior," *British Journal of Social Psychology*, vol. 40, no. 1, pp. 79–98, 2001.

[6] P.-R. Lei, T.-J. Shen, W.-C. Peng, and I.-J. Su, "Exploring spatial-temporal trajectory model for location prediction," *MDM*, 2011.

[7] V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang, "Collaborative location and activity recommendations with gps history data," *WWW*, 2010.

[8] H. Gao, J. Tang, and H. Liu, "gscorr: modeling geo-social correlations for new check-ins on location-based social networks," *CIKM*, 2012.

[9] K. Farrahi and D. Gatica-Perez, "Discovering routines from large-scale human locations using probabilistic topic models," *ACM Trans. Intell. Syst. Technol.*, 2011.

[10] E. H.-C. Lu, C.-Y. Lin, and V. S. Tseng, "Trip-mine: An efficient trip planning approach with travel time constraints," *MDM*, 2011.

[11] B. H. Sheppard, J. Hartwick, and P. R. Warshaw, "The theory of reasoned action: A meta-analysis of past research with recommendations for modifications and future research," *Journal of Consumer Research*, vol. 15, 1988.

[12] D. Hedeker, B. R. Flay, and J. Petraitis, "Estimating individual influences of behavioral intentions:an application of random-effects modeling to the theory of reasoned action," *Journal of Consulting and Clinical*, vol. 64, no. 1, pp. 109–12–, 1996.

[13] M. Fishbein and I. Ajzen, "Belief, attitude, intention, and behavior: An introduction to theory and research," *Book*, 1975.

[14] Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui, "Exploring millions of footprints in location sharing services," *ICWSM*, 2011.

[15] M. Kim and D. Kotz, "Extracting a mobility model from real user traces," *INFOCOM*, 2006.

[16] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, 2008.

[17] B. Jiang, J. Yin, and S. Zhao, "Characterizing human mobility patterns in a large street network," *PHYSICAL REVIEW*, 2009.

[18] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong, "Slaw: A new mobility model for human walks," *INFOCOM*, 2009.

[19] J. Lindqvist, J. Cranshaw, J. Wiese, J. Hong, and J. Zimmerman, "I'm the mayor of my house: Examining why people use foursquare - a social-driven location sharing application," *CHI*, 2011.

[20] H. Cramer, M. Rost, and L. E. Holmquist, "Proceedings of the 13th international conference on human computer interaction with mobile devices and services," *MobileHCI*, 2011.