What Cuisine Do You Like? - Improving Dining Preference Prediction Through Physical Social Locations

James She, Anna Vassilovski, Alvin Hon
HKUST-NIE Social Media Laboratory, Department of Electronic & Computer Engineering
Hong Kong University of Science & Technology Clear Water Bay, Kowloon, Hong Kong
{james.she, a.vass, kwhon}@ust.hk

***Abstract*—Social behaviors such as dining preferences are inextricably linked with physical social locations (e.g., home, work, and hangout location), rather than just due to the personal interests/cultures and influences from social peers. With the uses of location-based services in online social networks over smartphones, such physical social locations are easily available as an effective alternative to infer dining preferences. Results show that the prediction of individual dining preferences using physical social locations outperforms the common approaches simply using social information from peers.**

KEYWORDS

Social computing, physical social locations, inference.

I. INTRODUCTION

ONLINE social networks (OSN) over smartphones become a common medium for interactions between individuals, which are often used for sharing their daily social decisions among peers. These OSNs have even utilized the powerful GPS or advanced localization capabilities from smartphones to provide users location-based services to recommend things and lifestyle choices available at their current locations as showed in Fig. 1. Users in these OSNs generally share their lifestyle experiences not just through textual descriptions but also various multimedia such as the combinations of locations, audio, photos, maps, etc. Users became parts of an OSN by building a profile that desrcibes their background, including personal information and their approximated physical social locations (PSL), e.g., home, work and preferred hang out areas in OpenRice and Facebook in Fig. 2.

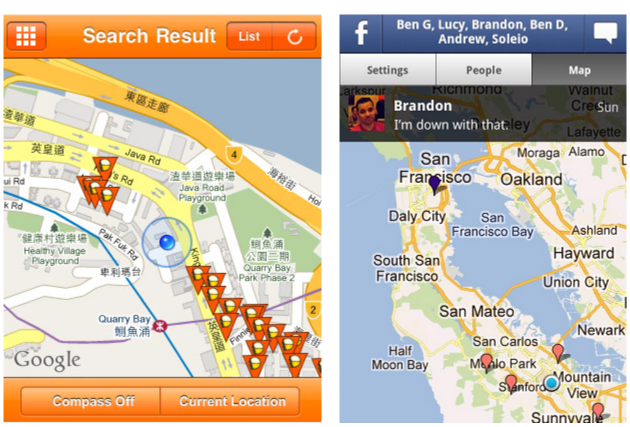

Fig. 1. Location information on mobility social network

Fig. 2. Physical social locations on a) OpenRice; b) Facebook

Many OSNs allow individuals to connect with their trusted peers or experts who share common interests or specialized expertise, in order to make better lifestyle choices from where to dine, and travel, what to buy, etc. All these decisions could be intuitively concluded as either or combined results of 1) the homophily factors like individual cultures or interests; 2) the social influences from peers such as recommendations from friends or critiques from trusted experts; or 3) the confounding factors due to the environment such as an individual work location. For the first two factors, related studies on inferring interest preferences on electronic resources are widely studied [1][2]. Some recent studies [3][4] investigate the relationships of some location histories and the preference prediction.

However, it is unclear how effective in preference prediction due to the confounding factor, such as PSL, when compared to the general inference approaches using peer information. A general approach inferring user dining preferences would use the peer preference information [5]. With additional information about users PSL easily available now via their personal profiles and social networking records in an OSN, which may offer an alternative approach and performance of preference inference.

This paper investigates how PSL of an individual could affect the prediction of his/her dining preferences through an OSN formed by a community of dinners and food lovers, e.g., OpenRice [6], Retty [7], Singapore Food Guide [8] in Asia and Urbanspoon in US and Canada [9].

This paper is organised as below. Section II describes the dataset in this work. The prediction model using the k-Nearest Neighbor (k-NN) supervised learning algorithm is discussed in Section III. The simulation results are discussed in Section IV. s Section V concludes the work with possible future works.

II. Dataset

To examine the potential of inferring user dining preference using PSL, OpenRice is identified as a good OSN candidate because it allows users with the ability to build a profile describing important locations in their daily life (such as home, work, and hangout locations). It also let users to build a peer network around their dining preferences by way of restaurant reviews, location information and a network of favorite users. It reviews over 29,000 restaurants in various districts of Hong Kong and categorizes restaurants by cuisine types.

When users build their personal profile on OpenRice, the OSN identify their home district, work district, favorite hangout districts, favorite dining cuisines, favorite dining dishes, and favorite dining restaurant atmosphere types. These preferences are based on categories provided in the OSN. Within the dataset collected from this OSN, there are dining and visiting histories on 64 districts, 38 cuisine categories, 33 dish categories, and 24 dining atmosphere types. On average, each user provided 3 to 12 profile features regarding their dining preferences and location based information.

Like in many OSNs, majority of users do not provide the complete information. However, there exists a ranking page in OpenRice listing the most active 500 users. Among these 500 users, the users who are missing to provide minimum PSL information will be eliminated. A dataset of around 400 users was compiled, including profile information, peer user information, and dining interests by way of visited restaurants. For these 400 users, there are over 10,000 location based records and over 12,000 dining interest records. The peer users of a particular user are identified by those whom he/she is followed as the favorite users. Within these 400 users, 60.8% of them provide hangout area information, 56.4% provide home area information and only 47.6% provide work area information. The purpose of collecting and identifying different PSL attributes is to investigate the ability of different PSL towards the inference performance for the dining preferences.

III. Dining Preference Prediction Model

In order to predict dining preference by way of user PSL, an inference model of user dining preference is developed. The model uses cuisine preference as a proxy for dining preference, which is generally specific to a particular restaurant atmosphere and embodies the types of dishes served at a particular restaurant location. The model adopts a k-NN algorithm to infer cuisine preference using available user PSL information. Algorithm 1 shows the pseudo-codes used for inference, which is commonly used in recommendation systems [10][11].

Algorithm 1 k-NN in predicting preference

Input: Parameter k, Training users $T_{training} = \{x_i, y_i\}, i = 1...n$, Test users $T_{test} = \{x_j, y_j\}, j = n+1...N$ with predicted preference $y_j = 0$

Output: Test users $T_{test} = \{x_j, y'_j\}, j = n+1...N$ with predicted preference values y_j'

for $each\ x_i\ in\ T_{test}$ **do**
  $initialize\ K\text{-}neighbours[]$
  for $each\ \{x_j, y_j\}\ in\ T_{training}$ **do**
    $s = S_J(x_i, x_j)$
    if $!K\text{-}neighbours.Isfull()$ **then**
      $K\text{-}neighbours.Pushback(\{s, y_j\})$
      $K\text{-}neighbours.Sort(Decreasing)$
    else if $s > K\text{-}neighbours[k].s$ **then**
      $K\text{-}neighbours.Popback()$
      $K\text{-}neighbours.Pushback(\{s, y_j\})$
      $K\text{-}neighbours.Sort(Decreasing)$
    end if
  end for
  y_i'$= Predict(K\text{-}neighbours[])$
end for

A. The Inference Model

The dataset can be expressed as $D = (x_i,\ y_i)$ for each user $i = 1, ..., n, n+1, ...N$, where N is the number of users and n is the number of users selected as a training set in the model. In this work, half of the data set (around 200 users) selected as the training data, and the remaining 200 users as the testing data. For user i, x_i represents the vector set of predicting factors for a given dining preference y_i taken from their profile and peer information.

Each feature of $(x_i,\ y_i)$ is modeled as an indicator variable. The indicator y_i, represents the presence or absence of a particular cuisine preference as +1, -1 respectively. The dataset contains 38 potential cuisine types for the prediction. For example, a home district predictor represents the vector of potential districts with a value of 1 for presence of the district and -1 for the absence of the district in the users home attribute. This configuration is employed on the dataset using various types of predictors, including PSL location only (work, home, hangout district), peer only (social neighbors, followee), and the combinations of them.

To infer of a cuisine type for a particular user i denoted as U_i, the k-NN algorithm is employed to find the k most similar users to user i. The parameter value k introduced in the algorithm is tuned to select the value that achieves best inference performance over the training set. The similarity measure $S_J(U_i, U_j)$ between user i and user j, is using Jaccard similarity over their predictor feature vector.

The k-NN for a target user i is done by identifying the k closest users in the training set, in terms of, inverse distance defined by Jaccard similarity. This similarity is the union of common attributes in the feature vectors for user i and user j as a fraction of the total length of the feature vector:

$$J(U_i, U_j) = \frac{x_i \cap x_j}{x_i \cup x_j} \tag{1}$$

The similarity $S_J(U_i, U_j)$ between user i and user j is:

$$S_J(U_i, U_j) = \frac{1}{1 + D_J(U_i, U_j)} \tag{2}$$

where

$$D_J(U_i, U_j) = 1 - J(U_i, U_j) = \frac{x_i \cup x_j - x_i \cap x_j}{x_i \cup x_j} \tag{3}$$

There are a number of similarity measures [12][13] employed in the social network literatures, e.g., Jaccard and cosine similarities. However, Jaccard similarity is a commonly used measure due to its simplicity. This is enough for this work, in which the feature vector represents indicator variables showing the presence/ absence of a particular feature value (i.e., dining preference). Once the distance between user i and every other user j in the training set is obtained, they are ranked in increasing order and the users corresponding to the top k smallest distances are identified to be the k nearest neighbors for the target user i. The inference step examines the value of the interest being inferred for each neighbor and identifies the prediction for the target user based on the weighted majority equation. The target user i is therefore predicted to have a cuisine interest, if the average of the k neighbors interest in the cuisine is non-negative:

$$Y = \frac{1}{|J|} \sum^{J} S_J \cdot y_j \tag{4}$$

where S_J is the similarity between the target user i and corresponding neighbor j. Here, $Y \subseteq$ [-1, 1] is the prediction value obtained by summing all y values with similarity S_J weighted over the neighbor set J.

B. Various Predicting Factors

The inference model of dining preference prediction is applied on the dataset to infer a cuisine interest based on: peer only information, location only information, and the combination of both.

Peer only information represents the favorite users followed by a target user. The dining preference prediction will be relying on the the preferred interests in peer network information available from the closest neighbors (i.e., the most favorites users). Such inference performance is used as a baseline for comparing the performance of the inference algorithm using various attributes of PSL. Location only information represents key individual locations identified by users in their profile regarding their home, work, and hangout locations. These are compared against the use of the peer network information and paired with the peer network information for a combined inference.

IV. Result and Discussion

The application of the inference model of dining preference prediction achieves high performance in recall when pairing individual location information with peer network information as compared to using peer network information alone in the cuisine inference.

The baseline predictor is chosen to be information from social peers but not recommendations learnt by personal interests or cultures, because the later one is obviously important factor for inference.

From the recall results in Fig. 3, it shows the performance of dining preference inference for each location introduced in the inference predictors in combination with the peer information. The inference was best using the combination of home and peer information followed by hangout location and peer information. Home information is likely a strong indicator for cuisine preference as compared to peer information as a target user may be interested in following peers with a similar locational environment. Moreover, by comparing the results between different attributes of PSL (i.e., home, work, and hangout), home and hangout areas achieve better results than using the work area as the indicator for cuisine preference prediction.

The inference using PSL motivates the use of easily available LBS data from OSN to offer an alternative and more efficient dining preference inference. With such approach, a restaurant is likely to be recommended to users who share similar PSL and possibly meet in person physically at the restaurant for more interesting and effective social networking.

With regard to the k-NN algorithm, various levels of recall are achieved for different k values as shown in Fig. 3. Values of k from 1 to 10 were tested in the analysis. Using recall as the metric for overall performance would support the use of k = 2 as a good parameter value for the k-NN algorithm.

V. Conclusion

Many existing OSNs infer user interests using available social peer information. However, for some social activities that occur in physical situation, users are interested in finding options that match their interests and lifestyles at their current physical locations which is not enough simply achieved by using social peer information. Using individual PSLs presents an alternative to improve the inference of user interest, especially for social network with LBS about dining preferences.

This paper has investigated the effectiveness between user dining preference and their PSLs. The results showed that there are improvement in prediction performance when the inference technique incorporate the use if individual PSL.

This work is currently being extended to evaluate other similar OSNs for dining recommendations in other countries. It will be interesting to see whether the performance gains also occur due to the individual PSL. In addition, many mobile social network applications are able to track the on-going user mobility. These data of user physical social trajectories may lead to a better inference for recommending more than just a suitable dining location, but a personalized routine

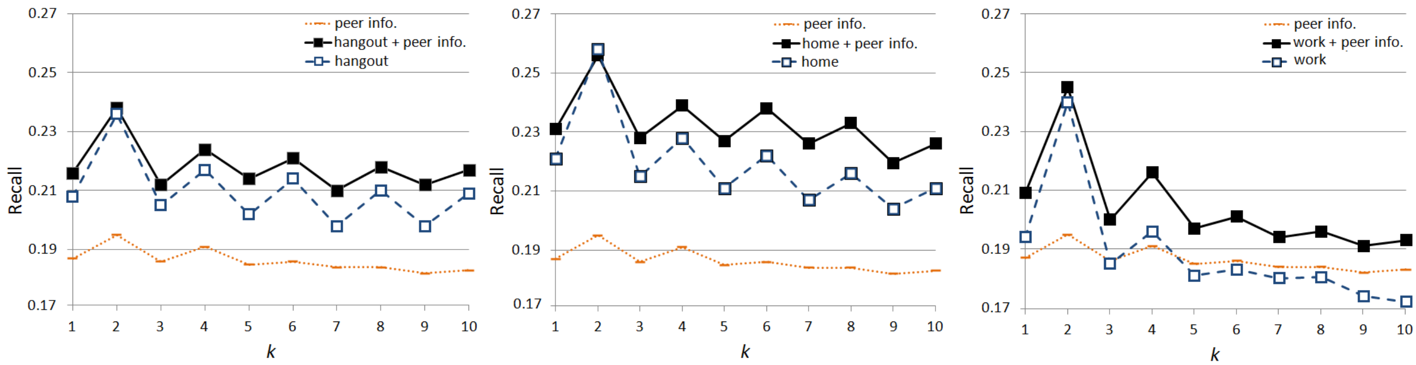

Fig. 3. Average recall at various levels of k under different combination of predictors

of multiple dining locations in the whole evening for better dining and travel experience. To the best of our knowledge, it is the first attempt to evaluate the inference performance subject to the confounding factors like PSL for better dining preference prediction, which is practical for possible real-world deployments in OSNs like OpenRice.

ACKNOWLEDGMENT

The authors would like to thank Ken Fong from HKUST-NIE Social Media Lab for experimental improvement.

REFERENCES

[1] Z. Wen., et al. On the quality of inferring interests from social neighbors. *Proc. of the 16th ACM SIGKDD Intl. Conf. on Knowledge discovery and data mining*, 2010.

[2] T. Hogg, Inferring preference correlations from social networks. *Electronic Commerce Research and Application*, 2010.

[3] A. Noulas., et al. Inferring Interests from Mobility and Social Interactions. *Proc. of Workshop on Analyzing Networks and Learning with Graphs (colocated with NIPS09)*. Whistler, Canada. Dec. 2009.

[4] Y. Zheng., et al. GeoLife2.0: A Location-Based Social Networking Service. *Mobile Data Management: Systems, Services and Middleware*, 2009.

[5] Y. Zheng., et al. Recommending friends and locations based on individual location history. *ACM Tran. on Web*, vol. 5, no. 1. Feb. 2011.

[6] OpenRice. http://www.openrice.com/

[7] Retty. http://retty.me/

[8] Singapore Food Guide. http://www.hungrygowhere.com/

[9] Urbanspoon. http://www.urbanspoon.com/

[10] Sun Bo., et al. Study on the Improvement of K-Nearest-Neighbor Algorithm. *Intl. Conf. on Artificial Intelligence and Computational Intelligence*. 2009

[11] X. Geng., et al. Query Dependent Ranking Using K-Nearest Neighbor. *Proc. of the 31st annual international ACM SIGIR conf. on Research and development in information retrieval*. 2008.

[12] A. Anagnostopoulos., et al. Influence and correlation in social networks. *In KDD '08*, pp. 7-15. 2008

[13] I. Guy., et al. Same places, same things, same people?: mining user similarity on social media. *Proc. of the 2010 Conference on Computer Supported Cooperative Work*, pp. 41-50 ACM. 2010.