Context-Aware Prediction of User’s First Click

Liang Wu †,‡,§  Alvin Chin †  Yuanchun Zhou †  Xia Wang †
Kangjian Meng †  Yonggang Guo †  Jianhui Li †

† Nokia Research Center
‡ Computer Network Information Center, Chinese Academy of Sciences
§ Graduate University of Chinese Academy of Sciences

{alvin.chin, xia.s.wang, kangjian.meng, yonggang.1.guo}@nokia.com
{wuliang, zyc, lijh}@cnic.cn

ABSTRACT
Location-based services has attracted attentions from both industry and academia. The development of position tracking technologies and the increasing popularity of smart phones has collected large amounts of contexts. An important issue is how to leverage the rich contexts to predict a user’s need accurately. In this paper, we propose a novel approach to predict the product type that a user will first click in an e-commerce application, after they update their location manually. Our proposed approach models the problem as a multi-label classification. We introduce three sets of features including location feature, time feature and behavioral feature. We use the Periodica algorithm [10], which was designed to mine the periodic behaviors of moving objects, to generate a series of periodicity templates. The templates are further exploited as behavioral features. Finally, we design several experiments using a real world dataset collected by an e-commerce application called WuXianGouXiang, which is developed by Nokia Research Center, Beijing and was released in September 2011. We have obtained a dataset from the service logs and the dataset contains over 3000 registered shops and 20000 users. Our experimental results demonstrate that the three sets of features contribute significantly to the classification of different users and the best result is achieved when using all of them.

Categories and Subject Descriptors
H.2.8 [Database Applications]: Learning
I.2.6 [Artificial Intelligence]: Learning

General Terms
Algorithms, Design, Experimentation

Keywords
User Behavior, Location Based Services, Classification, First Click Prediction

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ContextDD ’12, August 12, 2012, Beijing, China.
Copyright 2012 ACM 978-1-4503-1553-1/12/08 ...$15.00.

1. INTRODUCTION
With the recent advancement of pervasive and ubiquitous computing, location-based services has attracted more attention from both academia and industry. Applications like Foursquare and Yelp are very popular among users, while an application called Weixin has passed the 100 million users milestone in China. These applications provide interesting functions for their users in the cyber-physical realm that bridge the gap between physical and virtual environments, while at the same time collecting a history of their location and activity data. The large scale user generated information provides many monetization opportunities. As more contextual information like locations, user speed and check-in time can be collected by the mobile devices, finding a proper method for modeling user preference and recommending items according to certain contexts has been a key issue in mobile recommendations.

Traditional product recommendation algorithms recommend items by analyzing user preferences and product characteristics. By their different ways of modeling the relationship between user preference and items, state-of-the-art recommendation algorithms can be classified into three main types, collaborative filtering (CF) models [6, 15, 12], content-based models [14] and hybrid approach [3, 5]. CF-based models recommend items by exploring the users who bought similar products and/or the products which were bought by similar users. Content-based models have their roots in the information retrieval community and they recommend items by computing the similarity between products. Hybrid approach combines the previous two models.

Most algorithms used in traditional recommender systems including CF-based and content-based models do not take the contextual constraints of preferences into considerations. The scenario is more complicated when contexts including time and location are obtained. For example, it is proper to recommend a deal or discount on a restaurant which is close to the workplace at lunchtime on weekdays. But at the same time on weekends, the consumers may still be at home and sleeping, so invalid recommendations will be provided without considering the temporal constraints. Therefore to provide more accurate recommendations, a key issue is to predict a user’s need based on contextual information.

WuXianGouXiang is a context-aware product recommender application which is developed by Nokia Research Center,
Beijing. The application can provide nearby deals according to a user’s location and preference. When users need recommendations, they can update their positions and check available deals nearby. In this paper, we focus on predicting the product type of a user’s first click, as it can best describe the user’s need.

We model the type prediction problem as a multi-label classification task. A binary classifier is trained for each type of product. The data we have contains only the position including latitude and longitude, time and the product type. So, a fundamental issue is to extract effective features from them. Fortunately, a user’s first click behavior is not random and we find several interesting patterns from the dataset. We first exploit the time and location of the users’ logs as classification features. We use the Periodica [10] algorithm to generate several probability distribution templates by analyzing the logs. Two sets of behavioral features including the probability distribution templates are proposed to leverage the behavioral observations. The experimental results demonstrate that time, location and behavioral features are effective for the classifiers. These three sets of features contribute significantly to the classification of different users. The best result is achieved by using them all.

The main contributions of our paper lies in three aspects:

- **First Click Behavior Prediction**: We model the user’s first click behavior as a multi-label classification task and then design three sets of features including time, location and behavior to train the classifiers based on comprehensive observations.
- **User Behavior Modeling**: We design several behavioral features to describe the characteristics of consumers. We adopted a periodic behavior mining approach to leverage the periodic preferences. The features are proven to be useful for improving both precision and recall.
- **Real Evaluation**: We evaluate our method using a large-scale real life dataset obtained from a mobile application called WuXianGouXiang, which is developed by Nokia Research Center, Beijing and was released in 2011.

The remaining sections of the paper are organized as follows: In section 2, we present some related work on analyzing contextual user behavior and context-aware product recommendations. Section 3 describes the application, the dataset we use and the architecture of our proposed approach. The periodic behavior mining approach is discussed in section 4, while our proposed model is presented in section 5. In section 6, we present experiments to show the effectiveness of the proposed method. Section 7 carries out the discussions and finally, section 8 concludes the paper.

2. RELATED WORK

In this section, we will present some related work on product recommender systems and context-aware user behavior modeling.

2.1 Product Recommendation

Collaborative filtering is the most popular recommendation algorithm and can be divided into two parts, user-based CF [8] and item-based CF [16]. User-based CF algorithm computes a user’s preference based on what users have purchased in the past and thus recommending those new items which were bought by the users with a similar preference. Item-based CF algorithms quantify a product’s characteristics based on the group of users who purchased it. A challenging issue of CF-based models is how to conquer the cold start problem. As collaborative filtering is based on prior transaction histories, it is hard to model new users and new items. Content-based algorithm is another popular recommender model [14]. It tries to recommend items that are similar to those that a user has purchased before, and it can further exploit the profiles of products and consumers. Content-based recommendation approaches are usually applied as a make-up for the CF-based algorithms to solve the cold-start problem as they depend less on purchasing history.

It is important to consider the contextual information when recommending proper products in certain circumstances. For example, the recommendations for a traveler provided by a travel package recommender system can be very different in different seasons. In order to provide better recommendations, researchers have proposed to incorporate several contextual features into a learning to rank framework [2]. Yehuda further adopted temporal constraints into collaborative filtering [7]. Xiang et al. [20] proposed a Session-based Temporal Graph to model short-term and long-term preferences. Instead of recommending items, we focus on modeling by predicting a user’s first click based on a mobile application called WuXianGouXiang. Based on the prediction of the product type of a user’s first click behaviors, the recommender system can directly recommend what the user needs by analyzing the contexts.

2.2 User Behavior Modeling

A fundamental issue for providing personalized services, i.e. product recommendation and advertising according to a user’s preference and contexts, is to identify the characteristics and differences of user behaviors. Both in industry and academia, many works are focused on leveraging historic user behavior to provide more accurate advertising. Behavioral Targeting has been proven to be useful for improving business campaigns [21, 9]. For example, Cao et al. [4] proposed an association rules mining-based approach for mining user habits on mobile phones and further discovered similar users by using a factorization model to extract the common latent user habits.

Wang et al. [19] proposed to predict and identify user behaviors with temporal constraints. Since our scenario is different from theirs, i.e. the behavior we plan to model is easy to extract from the logs, this paper is only focused on predicting the product type of a user’s first click after they update their positions by mining the contexts, which includes temporal constraints, location feature and user behavioral preference.

3. METHOD AND APPROACH

In this section, we present the mobile application and the architecture of our approach.
3.1 WuXianGouXiang

In this paper, we propose to predict the product type that the user first clicks after they update their position. The experiment is based on a real life application called WuXianGouXiang which provides nearby discounts and coupons according to user preference and location. Users can open the application and update their position if they need recommendations.

WuXianGouXiang is a mobile application developed by Nokia Research Center, Beijing and was released in September, 2011. WuXianGouXiang is based on the O2O (online to offline) business model, which promotes the physical commerce by online services and recommendations. Unlike other O2O applications, e.g., Groupon and Fanfou, which provide deal-of-the-day on their web sites, WuXianGouXiang is based on exploiting context data from smart phones, in particular, location data using GPS and cell ID. Users can select the product type and locations they are interested in. Thus, it can recommend deals to users.

There are five types of coupons and discounts offered in WuXianGouXiang including food, clothes, hotel, entertainment and daily necessities. Every product belongs to one of the five types. Figure 1 shows the user interface of WuXianGouXiang, which displays the deals in Beijing.

In this paper, we propose to predict a user’s first click behavior after they update their location. When a user needs recommendations, she tends to click the product which she is interested in most, so the first click behavior can best describe a user’s need since she may have a look at other products after she finds what she likes best. In WuXianGouXiang each shop offers some new campaigns every week, so the number of products is quite huge and the interaction between users and products is very sparse. In order to solve the sparseness problem and better understand the relation-ship between the need and context of users, we predict the type of the products instead, as we believe that this can better represent a user’s need and the data is much more dense.

3.2 Architecture

Figure 2 displays the system architecture of our proposed approach. The context logs contain the time and location where a user uses the application and the product type of the first click behavior. The time is first directly exploited as time features. By observing the time distribution of each user in the dataset, two user behavioral preferences are found. Based on the observations, we propose two kinds of behavioral features. Since the dataset we use has over 3000 shops and 20000 users, the positions of users hardly overlap precisely, i.e. the raw GPS points are quite sparse. So, we first cluster these GPS points using a density-based clustering approach, namely DBSCAN [13]. Each cluster is labeled by the product type which was most viewed by users in this cluster. Then, the cluster ID and product type are used as location features.

After extracting all the features, two classification models, Support Vector Machine and K-Nearest Neighbor are trained to predict a user’s first click behavior.

4. PERIODIC BEHAVIOR MINING

As mentioned earlier, the distribution of click type follows a periodic manner. Figure 3 shows the product type distribution of a user’s first click behavior on Monday, Tuesday, Saturday and Sunday. It can be seen that the distributions of Tuesday and Saturday are quite similar. Therefore, our goal is to leverage the periodic manner to improve the accuracy of the prediction.

The Periodica algorithm was proposed by Zhenhui Li et al. in [10, 11]. The algorithm first collects all the reference spots by clustering the raw GPS points. Based on the reference spots, Discrete Fourier Transform and autocorrelation are combined to detect all the possible periods including long-term periods and short-term periods [18]. By collecting all the segments that share a same period, Periodica then produces a generalized probability distribution for each periodic behavior by hierarchically clustering the segments.
In our scenario, the periods are much simpler. Unlike the trajectory information used in [10, 11], which tracks all the time while the objects move, our application records a user’s location only when they use the application and upload their position. However, noisy and useless short-term periods may be detected if we run the algorithm straightforwardly. On the other hand, since our service has been run for about a year, it is not feasible to find long-term periods. Thus, we set the periods as 24 hours by hand for simplicity. That is, we ignore those long-term and short-term periods which have a period longer or shorter than one day.

Finally, the Periodica algorithm uses a probability model to characterize the periodic behaviors. Each periodic behavior is modeled as a generalized probability distribution template through hierarchically clustering all the sequences belonging to it. As we define the periods as 24 hours in this paper, our objective is to cluster all the “days”. By averaging the probability distribution of “days” of a same cluster, a probability distribution template is produced for each cluster. Li et al. [10] define the periodic behavior distribution template as a Categorical Distribution Matrix, here we use similar definitions which we explain below.

**DEFINITION 1.** Let \( T = \{ t_1, t_2, ..., t_r \} \) be a set of relative timestamps (here the timestamp ranges from 0 to 23 representing the 24 hour period), and \( x_k \) be the product type variable indicating the product type at timestamp \( t_k \). \( P = [p_1, ..., p_T] \) is the probability distribution template with each column \( p_k = [p(x_1 = 1), p(x_2 = 2), ..., p(x_k = m)] \) being an independent user preference distribution vector satisfying \( \sum_{i=1}^{m} p(x_i = i) = 1 \), where \( m \) is the number of product types and equals five in this paper.

As for the clustering part, Li et al. [10] adopted an agglomerative hierarchical clustering model to cluster the segments and used the well-known Kullback-Leibler divergence as the distance measure. Kullback-Leibler divergence is also known as relative entropy, which calculates the distance between two probability distribution matrices \( P \) and \( Q \):

\[
KL(P||Q) = \sum_{k=0}^{23} \sum_{i=1}^{5} p(x_k = i) \log \frac{p(x_k = i)}{q(x_k = i)}
\]

If \( P \) and \( Q \) are similar, \( KL(P||Q) \) is small. KL divergence is a non-symmetric measure and becomes infinite when any entry of either matrices contains zero. So, we use Add-One smoothing to avoid this problem, that is, the visit number of every product type is added by one.

Finally, we get several periodic behavior templates by clustering all the “days”. When predicting the product type of a user’s first click, we first construct a product type probability distribution of the user’s historic first click behavior and then calculate the KL divergence between the user and each periodic behavior template. According to the given timestamp, the corresponding column of the template is extracted as the classification features, e.g., if the given time is 17:25, \( p(x_1 = 1), p(x_1 = 2), ..., p(x_1 = 5) \) are used as behavioral features.

### 5. FIRST CLICK BEHAVIOR PREDICTION

As we model the product type prediction problem as a multi-label classification, an important issue is to extract effective features from the contexts. In this section, the proposed model and three sets of features will be introduced in detail.

#### 5.1 Multi-Label Classification

Given a user’s context, our objective is to predict the product type the user will first view. Specifically for our WuXiGouXiang application, the problem can be modeled as a classification with five labels, i.e. food, clothes, hotel, entertainment and daily necessities. A commonly used way is to transform the multi-label classification into binary classification [17, 22] where a binary classifier is trained for each product type. As every first click behavior contains only one type, the product type with the highest prediction confidence will be selected.

We use two popular classification models, Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) for our problem. The models are both implemented by a Java-based machine learning library named JavaML [1].

#### 5.2 Location Clustering

In this section, we discuss the proposed approach of finding areas by clustering the raw GPS points. As many behaviors are random, the clustering results may be influenced by the noise considerably if we use an algorithm which cannot handle noisy data like the K-Means algorithm.

Thus we adopt the DBSCAN algorithm [13] to find the periodicity in user behavior. DBSCAN does well in filtering out the points with lower density. In the density-based clustering theory, the density is obtained by counting the number of points in a region of a predefined radius, called Eps a region is positive if the density is over a predefined threshold, MinPts. The positive areas are classified as clusters while the rest are filtered out. The most difficult task of applying DBSCAN is to set proper parameters. We set MinPts as 50 and the radius Eps as 50 meters by hand based on our observations, as the visiting times of popular Point Of Interests (POI) is usually very dense and several adjacent POIs may be grouped together mistakenly if the Eps is enlarged.

![Figure 3: An Example of a Product Type Distribution of a User's First Click Behavior for Monday, Tuesday, Saturday and Sunday.](image-url)
Figure 4: Snapshot of Clustering GPS Coordinates in Beijing.

Figure 5: User Context Log Example.

Figure 4 presents a snapshot of the clustering result of the GPS points of the 3rd Ring Road in Beijing, China. The highlighted flags are the cluster centers, where the number represents the cluster ID. Furthermore, we assign a product type to each cluster. Each cluster contains a lot of user behavior logs, we select the most viewed product type of each cluster as its tag. The tag is further exploited as the location feature, i.e. when a new instance is input to be predicted, if its position to the nearest cluster is within a certain threshold, then the assigned product type of the location will be listed as a location feature. Our intuition is that customers tend to check similar products which the location is well known for.

5.3 Behavior Analysis

5.3.1 Switching Preference

We propose two kinds of behavioral features based on a user’s clicking preference. Figure 5 displays a sample log of two users. It can be seen that the first user with ID 12632 stays on one product type for several clicks, while the other user with ID 12628 switches between different types quite frequently. In order to quantitatively analyze the phenomena, we study the diversity of stay times of users by computing the entropy of product types (food, clothes, hotel, entertainment and daily necessities). The result is shown in Figure 6. Smaller entropy indicates that users stay at one product type for similar times. From the figure, we observe that over 80% of users have their entropies smaller than 0.75 and about 90% of users have their entropies smaller than 1. Thus, we can say that most users have a stable preference on switching between product types when using the application. We name this behavioral characteristic as switching preference. Further, we design some features based on this observation as will be explained in Section 5.4.

5.3.2 Periodic Preference

As discussed in Section 4, the product type distribution is relatively periodic. Based on the results of Section 4, we use the probability distribution template as periodic features. That is, when given an instance to predict, the most similar template will be selected compared with the user’s historic behavior. The column of the template with the same timestamp as the input instance is extracted as the periodic features.

5.4 Features

From our dataset and description above, we collect a total of 15 features. The 15 features belong to three categories, i.e. time features, location features and behavioral features.

- **Time Features**: As users tend to view the same type of products at similar times, e.g., people usually update their positions and check the coupons and deals of nearby restaurants at noon, we characterize time features in a scale of week and day.
  - hour_of_day: The hour of the day, which ranges from 0 to 23.
  - day_of_week: The day of the week, which ranges from 0 to 6.
  - is_holiday: If the date is a holiday, this feature will be set as "true", else "false".

- **Location Features**: As users tend to check the same type of products in similar places (e.g., users usually try to find a bar in Houhai, but try to find a good hotel when they are near the airport), we incorporate the
position as location features. Based on the previous discussion, we propose four location features.

- **latitude**: The latitude of a user’s position in the context log.
- **longitude**: The longitude of a user’s position in the context log.
- **cluster_id**: The id of the nearest cluster of which the distance is within the threshold.
- **type_by_frequency**: The product type which is most viewed by users in the location.

*Behavioral Features*: As users have a relatively stable preference on switching between product types and the product viewing behaviors of most users share a periodic pattern, from the results of Section 4, we design eight features of user behaviors grouped into switching and periodic preferences:

**Switching Preference**

- **previous_type**: The product type selected from a user’s previous click.
- **previous_type_has_lasted**: The number of stay times on the previous product type.
- **average_stay_times**: The average number of times a user stays on one product type.

**Periodic Preference**

- **periodic_probability_food**: The probability of clicking the food product type at this time on the template.
- **periodic_probability_clothes**: The probability of clicking the clothes product type at this time on the template.
- **periodic_probability_hotel**: The probability of clicking the hotel product type at this time on the template.
- **periodic_probability_entertainment**: The probability of clicking the entertainment product type at this time on the template.
- **periodic_probability_daily_necessities**: The probability of clicking the daily necessities product type at this time on the template.

## 6. EXPERIMENTS

In this section, we describe the dataset we use and experiment configurations. The experimental results and the differences of the three sets of features are also discussed.

### 6.1 Dataset

We conducted experiments with the user logs of WuXian-GouXiang from September 2011 to March 2012. The application offers nearby coupons and discounts based on a user’s location. Five types of deals nearby including food, clothes, hotel, games and daily necessities are offered in our application. The first click behavior after users update their positions is extracted in this paper as it can be the most informative behavior. The application records the position of users in a context log in the following format: latitude and longitude, or cell ID. For simplicity, we only choose the logs with a GPS position for analysis. The final dataset contains 262 users and 3231 first click behavior logs. Table 1 gives the detailed statistics.

### 6.2 Experiment Settings

10-fold cross validations are performed in our experiments: the logs are randomly partitioned into 10 subsamples and the validation is repeated 10 times. A single subsample is retained as the testing set and the remaining 9 subsamples are used as training data for each round, the ten results are averaged to produce the experimental results. The location feature and behavioral features including the periodicity templates are generated by using the training data in each round. Two popular classification algorithms, namely Support Vector Machine(SVM) and K-Nearest Neighbor(KNN) are selected to test the features. The SVM and KNN models are implemented by a Java-based machine learning library called JavaML[1]. We describe the features below:

1. **Time and Location Features**: In order to validate the effectiveness of our proposed behavioral features, we merge time features and location features into a group as the baseline. These features are all directly extracted from the context logs and thus are less dependent on user history.

2. **Behavioral Features**: This group contains the behavioral features generated by the switching preference and periodic preference.

3. **All Features**: This combines both time and location features and behavioral features.

### Table 1: Statistics for the Dataset

<table>
<thead>
<tr>
<th>Category</th>
<th>Number of Logs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>1518</td>
</tr>
<tr>
<td>Clothes</td>
<td>139</td>
</tr>
<tr>
<td>Hotel</td>
<td>29</td>
</tr>
<tr>
<td>Entertainment</td>
<td>149</td>
</tr>
<tr>
<td>Daily Necessities</td>
<td>1396</td>
</tr>
</tbody>
</table>

### Table 2: Performance of each type(SVM)

<table>
<thead>
<tr>
<th>Category</th>
<th>Number Of Logs</th>
<th>Prec.</th>
<th>Recall</th>
<th>F-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>1518</td>
<td>59.4%</td>
<td>71.1%</td>
<td>64.7%</td>
</tr>
<tr>
<td>Clothes</td>
<td>139</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Hotel</td>
<td>29</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Enter.</td>
<td>149</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Necessities</td>
<td>1396</td>
<td>60.4%</td>
<td>61.1%</td>
<td>60.8%</td>
</tr>
<tr>
<td>AVG</td>
<td>3231</td>
<td>54.0%</td>
<td>59.8%</td>
<td>56.8%</td>
</tr>
</tbody>
</table>

### Table 3: Performance of each type(KNN)

<table>
<thead>
<tr>
<th>Category</th>
<th>#Instances</th>
<th>Prec.</th>
<th>Recall</th>
<th>F-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>1518</td>
<td>56.3%</td>
<td>67.3%</td>
<td>61.3%</td>
</tr>
<tr>
<td>Clothes</td>
<td>139</td>
<td>6.5%</td>
<td>2.2%</td>
<td>3.2%</td>
</tr>
<tr>
<td>Hotel</td>
<td>29</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Enter.</td>
<td>149</td>
<td>4.4%</td>
<td>1.3%</td>
<td>2.1%</td>
</tr>
<tr>
<td>Necessities</td>
<td>1396</td>
<td>57.0%</td>
<td>54.1%</td>
<td>55.5%</td>
</tr>
<tr>
<td>AVG</td>
<td>3231</td>
<td>51.5%</td>
<td>55.1%</td>
<td>53.3%</td>
</tr>
</tbody>
</table>
Figure 7: F-Measure of Different Feature Set Using KNN

Figure 8: Micro-Average of F-Measures Using KNN

Table 2 and Table 3 show the experimental results by using SVM and KNN respectively with all the features. It can be seen that SVM outperforms KNN on average and they are not effective enough to predict the rare types: Clothes, Hotel and Entertainment. The reason may be that the proposed features including Time and Location features and Behavioral features are not effective enough to retrieve the types which appears too few times in the dataset.

Figure 7 displays the capability of different type of features when using KNN, it can be seen that behavioral features perform better than time and location features when experimented individually. In Figure 8, the blue line shows the micro average of the three sets of features. The micro average is an average over instances, i.e. the product type which has more instances contribute a larger weight to the final average value. It can be seen that the best result is achieved when using all the features.

In order to discover the different influences of the time and location features and behavioral features, we further perform an experiment and the result is shown in Figure 9. SVM is used in this experiment.

We define the activeness of the user with most first click behaviors as 1, then all other users’ activeness can be calculated by comparing with the most active user. In Figure 9, larger user activeness means that the user has more first click behavior logs. Firstly, when the experiment is based on users who have less history, Time and Location features outperform Behavioral features. However, as the users with more first click behaviors are added into the dataset, the performance of Behavioral features improves faster than Time and Location features. The experimental results prove that different features contribute significantly in different aspects to the classification of different users. As Behavioral features are mined from logs, the features are more effective if given more history logs.

In conclusion, behavioral features are better at exploiting the user behavior history, while Time and Location features are less dependent on historic data since they are designed to model the characteristics of the locations and times. So time and location features may be more useful for solving the cold-start problem if the user to be predicted has few logs.

7. DISCUSSION

As displayed in the experimental results, the proposed model is not effective enough for those product types which appear less times in the dataset. The reason is that the multi-label classification model sacrifices the accuracy on the rare types to improve the overall performance. However, predicting those rare product types is important for modeling a user’s need and recommending the right items. In future work, we will explore to better model the rare product types like hotel and entertainment by transferring knowledge from other types.

In real world applications, there will be a lot of monetization opportunities if users’ need can be accurately estimated. For example, based on the product type prediction of a user’s first click behaviors, we can improve the ranking of products in WuXianGouXiang. Based on the analysis of users’ first click behaviors in different areas and times, we can provide suggestions for commercial companies, e.g., where to open up a new store and when to perform a campaign.
8. CONCLUSION AND FUTURE WORK

In this paper, we proposed a novel classification-based approach to predict the product type of a user’s first click after they update their position. The contributions of our work are the following. First, we model the user’s first click behavior as a multi-label classification task and then design three sets of features including time, location and behavior to train the classifiers based on comprehensive observations. Second, we design several behavioral features to describe the characteristics of consumers. We adopted a periodic behavior mining approach to leverage the periodic behaviors. The features are proven to be useful for improving both precision and recall. Third, we evaluate our method using large-scale real life data collected by an application called WuXianGouXiang. The experimental results demonstrate that time, location and behavioral features are effective for the classifiers. The three sets of features are all important aspects as they contribute significantly in unique aspects to the classification of different users. The best result is achieved by using all the features.

In the future, we will try to improve the performance on rare product types of our model and validate our method with other datasets to see the effectiveness of our approach. In addition, as traditional recommendation algorithms cannot fully exploit the rich contexts, the first click prediction model can be used to implement a context-aware recommender system.

9. ACKNOWLEDGMENTS

We are very grateful to the WuXianGouXiang project team, Hao Yang, Minggang Wang, Shenhua Wang, Ke Zhang and Olivia Li. They designed and developed the service from scratch and allowed us to get access to the data and conduct this research work. Alvin Chin, Yuanchun Zhou, Xia Hao Yang, Minggang Wang, Shenhua Wang, Ke Zhang and Olivia Li are the contact authors. The work is partially supported by the Natural Science Foundation of China(NSFC) under Grant No. 61003138.

10. REFERENCES