Learning User Behaviors for Advertisements Click Prediction

Chieh-Jen Wang and Hsin-Hsi Chen
Department of Computer Science and Information Engineering
National Taiwan University
Taipei 106, Taiwan
{d96002, hhchen}@csie.ntu.edu.tw

ABSTRACT
Predicting potential advertisement clicks of users are important for advertisement recommendation, advertisement placement, presentation pricing, and so on. In this paper, several machine learning algorithms including conditional random fields (CRF), support vector machines (SVM), decision tree (DT) and back-propagation neural networks (BPN) are developed to learn user's click behaviors from advertisement search and click logs. In addition, four levels of features are extracted to represent user search and click intents. Given a user's search session and a query, machine learning algorithms along with different features are proposed to predict if the user will click advertisements displayed for the query. We further study the impact of feature selection algorithms on the prediction models. Random subspace (RS), F-score (FS) and information gain (IG) are employed to search for a predictive subset of features. The experiments show that CRF model with the random subspace feature selection algorithm achieves the best performance.

Categories and Subject Descriptors
H.3.3 [**Information Systems**]: *Information Storage and Retrieval.*

General Terms
Design, Experimentation, Measurement

Keywords
Advertisement click prediction, advertisement logs analysis, intent features representation.

1. INTRODUCTION
The commercial value of advertisements on the web depends on whether users click on the advertisements. Issues such as users' intent analysis, advertisement selection, and so on may affect the click probability of advertisements. This paper aims to design an intent based model for prediction of advertisement clicks. When a user submits a query, the model will predict whether at least one advertisement will be clicked before target advertisements are displayed. An essential part of the model is the extraction of intent-related features from a large-scale advertisement search and click logs. The research issues are what representations are more suitable to represent users' intents, how long the historical information is considered, and what machine learning algorithms and feature selection algorithms affect the prediction.

Four levels of features are extracted from the logs, including current impression level, the first impression level, the previous *n* impression level and the contextual impression level. Features of current impression level cover the current query, query terms overlaps, and more sophisticated semantic similarities between current query and previous query in a session. Features of the first impression level describe the initial goal of a session. Features of the previous *n* impression level are related to the historical information of the previous *n* queries. Features of the contextual impression level give access to all contextual information to the first query in a session. The intent may make clearer due to intent disambiguation through the contextual information.

Moreover, we investigate which features have more effects on predictions. In other words, we would like to know whether using all of the features is more accurate in predicting advertisement clicks of users or using only a subset of features selected by feature selection algorithms is still workable. Conditional random fields (CRF), support vector machines (SVM), decision tree (DT) and back-propagation neural networks (BPN) with different intent representations are explored. Random subspace (RS), F-score (FS) and information gain (IG) are employed to select the best feature sets. Finally, all models are tested on a large-scale advertisement search and click logs. The experimental results show that user's click behaviors are significant with regard to their search intents. The intent based CRF model with the RS feature selection algorithm performs the best.

The rest of this paper is organized as follows. The related work is presented and compared in Section 2. The experimental dataset used in this study is described in Section 3. The intent-related features in our work are introduced in Section 4. Experimental results are shown and discussed in Section 5. Finally, Section 6 concludes the remarks.

2. RELATED WORK
The researches of advertisements on the Internet have attracted significant attention recently [23]. To maximize the commercial value of advertisements, selecting the most relevant target advertisements to match users' intent best is crucial. Issues which need to be dealt with include (1) matching relevant advertisements for a query [1], (2) ranking of the candidate advertisements [16], (3) deciding how to display the advertisements on the search result page [9], (4) click prediction and analysis for the presenting advertisements [8], and (5) pricing of the advertisements [3].

Several machine learning algorithms such as logistic regression [22], maximum entropy [8], support vector machines (SVM) [6] and conditional random field (CRF) [11] have been adopted to predict the clicks of advertisements presented for a query. The systems incorporate various features extracted from multiple resources. Text features are based on information from bid term, content of advertisement itself and related advertisements [22]. Besides text features, demographics features based on user information such as age, gender, interests, occupation, and so on have been shown to be very useful for predicting advertisements click after online and offline testing [8]. Furthermore, mouse trajectory features extracted from a library search engine showed that fine-grained interactions have a high correlation to the predicted advertisement clicks [11].

Search intents have been studied at various levels of granularity and applied to different applications. At the query level, Broder [5] divided query intent into navigational, informational and transactional types. Nguyen and Kan [19] characterized queries along four general facets of ambiguity, authority, temporal sensitivity and spatial sensitivity. Manshadi and Li [18] classified queries into finer categories. At the session level, Radlinski and Joachims [21] mined intent from query chains and used it for learning to rank algorithms. Boldi *et al.* [4] created graphs with query phrase nodes and used them for query recommendation. Ashkan *et al.* [2] classified queries into two intent types such as noncommercial/commercial and navigational/informational, and trained a decision tree for classification. Guo and Agichtein [10] mined *research* or *purchase* intent from a search task, and applied the experiences in information access. Moreover, Guo and Agichtein [11] extracted groups of features such as Query, SERP Content, Result Quality, Interaction, Click and Context from a library search engine. These features are used to predict *receptive* and *not receptive* within the current session.

Feature selection chooses a significant subset of features from a dataset and reduces the number of features in machine learning. It aims to be both effective and efficient for machine learning processing. Feature selection algorithms can be classified into two categories: the filter approach and the wrapper approach [15][17]. The former selects important features first and then classifier applied them for classification. In contrast, the latter combines classifier with other optimization algorithms to perform feature selection.

These previous works consider information from queries and content of advertisements on advertisement click prediction, but do not capture the intent of users completely. Different from their work, this paper employs cues from users' intent and click behaviors to predict the advertisement clicks in the future. We explore users' historical information to gain more accurate prediction, adopt feature selection algorithms to improve the performance of prediction, and compare the importance of individual features.

3. ADVERTISEMENT CLICK LOGS

Microsoft AdCenter logs (abbreviated as AdLogs) are adopted to learn users' intent and advertisements click behaviors. The logs consist of 101 million impressions and 7.82 million clicks during 84 days from 10th Aug to 1st Nov 2007. An *impression* is defined as a single search result page described by a set of attributes. For each impression, its query terms, query time, page number of the results, number of advertisements displayed in results of this impression, number of advertisements displayed in previous pages for this query, reported advertisements ID, location ID and the associated session are recorded in the logs. A session is defined by a repeated search engine usage of intervals of 10 minutes or less, with a total session no longer than 8 hours. A *click* is defined as an advertisement click associated with an impression. For each click, the clicked URL domain name, click time, search result rank of the advertisements and the associated impression are recorded in the logs. In total, there are 40.6 million sessions and 5.06 million sessions contain at least one advertisement click. For each impression, the number of displayed advertisement varies from 0 to 8. The advertisement click through ratio (CTR) is defined as the number of impressions containing advertisement clicks divided by total impressions. Figure 1 shows CTR according to the number of advertisements displayed.

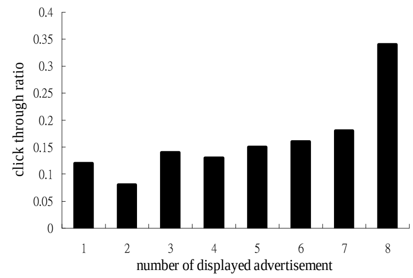

Figure 1 Average Click Ratio for Displayed Advertisements

For the purposes of promotions, some specific queries are issued or advertisements are clicked many times in a very short period of time by software robots. We follow some ideas proposed by Buehrer, et al. [7] to detect software robots in the web search. The following 3 criteria are used to purify the robot activities from the AdLogs: (1) issue queries more than 7 times in any 10 second interval; (2) issue queries at two distinct places at the same time; (3) click an advertisement more than one time in any 5 second interval. Moreover, some impression IDs are duplicated for some unknown reasons. We filter out those sessions, including robot activities or duplicate impression IDs from the AdLogs. Finally, 36.39 million sessions remain and 4.55 million sessions contain advertisement clicks among them. We further partition the sessions which contain at least one advertisement click on the first 56 days for training. Among the 3.12 million training sessions, 6.92 and 3.75 million impressions belong to non-click and click types, respectively. The sessions in the last 28 days are used for testing. Table 1 shows the details of the datasets for the latter experiments.

Table 1. Experiment Datasets

	Training	Testing
#sessions (clicks)	3.12M	1.42M
#sessions (non-clicks)	0	10.61M
#click impressions	3.75M	1.73M
#non-click impressions	6.92M	37.41M

4. FEATURE REPRESENTATION

Every impression q_i ($1 \leq i \leq n$) in a training session $s = (q_1, q_2, ..., q_{(i-1)}, q_i, q_{(i+1)}, ..., q_n)$ is represented as a feature vector and a labeled type. We label an impression *click* if it contains at least one advertisement click. Otherwise it is labeled *non-click*. Table 2 lists features at different feature levels and a description of each feature. Features capture information from q_i itself (Current Impression Level), the first impression q_1 (First Impression Level), the previous n impression $q_{(i-n)}$ (Previous n Impression Level), and all the contextual impressions q_1, q_2, ..., $q_{(i-1)}$ in s (Contextual Impression Level). The detail is described as follows.

Current Impression Level: These features aim to capture the query similarity, users' intent and location information from the view of the current query. They cover its position in a session, the length of the query terms in words, query terms itself, DMA level location ID from the issued query, cosine similarity and overlapping of query terms. The cosine similarity between q_i and $q_{(i-1)}$ is defined below.

$$Sim_{COS}(q_i, q_{i-1}) = \frac{q_i \bullet q_{i-1}}{|q_i| \times |q_{i-1}|}$$

The overlapping between q_i and $q_{(i-1)}$ is defined as follows.

$$Overlapping(q_i, q_{i-1}) = \frac{q_i \bullet q_{i-1}}{|q_i|}$$

In addition, we adopt 14 categories (exclusive of "Regional" and "World") on the 2nd level of the Open Directory Project (ODP) ontology to represent query categories. We look up ODP with a query and determine the weights of query categories by the distribution of returned web pages. Besides, 4,020 intent clusters are learned from MSN Search Query Log excerpt - RFP 2006 dataset [3] in our previous work [24]. Each intent cluster contains sessions of the same intent. Query category is specified by the distribution of the top 100 similar intent clusters in 14 intent types. Tree types (e.g., information, navigation, or transaction) of queries are determined by rules proposed by Jansen *et al* [13].

First Impression Level: These features aim to capture an initial search goal of a session. The features include the first submitted query terms in a session and time duration from the first query to the current query in a session.

Previous *n* Impression Level: These features aim to capture the advertisements clicks information of the previous n impression, including posted advertisements, clicked advertisements, their URL domain names along with ODP categories, intent types, time duration to current query and overlapping advertisements. In our experiments, n is set to 1 and 2. Similarly, we consult ODP to determine the categories of an advertisement. If not found, equal weight is assigned to all categories of the advertisement. As mentioned before, intent types of the clicked advertisements are identified by the distribution of the top 100 similar intent clusters.

Contextual Impression Level: The access history (q_1, q_2, ..., $q_{(i-1)}$) is also explored and encoded in the feature vector of q_i. These features represent a sequence of users' behaviors such as the number of advertisement clicks continuously, the number of continuous impressions containing clicked advertisements, and the nearest impression containing clicked advertisements. Additionally, click through ratio (CRT) with/without considering ranks of result pages are calculated. We determine the weight of intent types of submitted queries (CTQIntent) and clicked advertisements (CTAdIntent) in the access history as follows.

$$WI_m = \prod_{j=1}^{i-1} P_m(w_j \mid intent\ cluster\ set)$$

$$= \prod_{j=1}^{i-1} [(1-\lambda)P_m(w_j \mid intent\ cluster\ set) + \lambda P(w_j \mid AdLogs)]$$

Where WI_m denotes the weight of the type m intent, and P_m is a probability of the type m intent calculated by the distribution of the top 100 similar intent clusters. The weight of the feature is multiplied from q_1 to $q_{(i-1)}$. Here, w_j denotes a query or a clicked advertisement in q_j which is the impression at position j of the access history. However, the probability may be zero because no intent clusters are similar. We employ Jelinek-mercer method [14] which involves a linear interpolation of the collection model for smoothing. In our experiments, λ is set to 0.1. Moreover, we use the same method to deal with the weight of ODP categories (CTAdC) and the function is defined as follows.

$$WODP_m = \prod_{j=1}^{i-1} P_m(w_j \mid ODP)$$

$$= \prod_{j=1}^{i-1} [(1-\lambda)P_m(w_j \mid ODP) + \lambda P(w_j \mid AdLogs)]$$

For identifying clearer intent of clicked advertisements, we introduce a strategy for disambiguation. We postulate that the sequence of clicked advertisements in a session satisfies a user's intent. Thus, the clicked advertisements are coherent and co-related. The clicked advertisements surrounding a specific clicked advertisement form its context. The contextual information is employed to disambiguate the ODP categories of clicked advertisements. For an advertisement, we consult the ODP to collect all possible paths by its URL domain name. A *path* is defined as an ordered hierarchical structure of hyperlink labels from the root category to a leaf in the ODP [20]. Because a clicked advertisement may belong to more than one path, we must find its correct meaning. The goal of category disambiguation is to find the most semantically-coherent path of each advertisement. Take Figure 2 as an example. There are four possible trails which consist of probable paths from each clicked advertisement. The weight of a trail is the sum of similarity between this path and the other paths in the access history. The similarity of two paths is the number of common categories between these two paths. The number of common categories among paths reflects the degree of intent coherency. Therefore, a trail with the highest weight will be selected, and the trail contains the disambiguated path for the clicked advertisements to represent their intent in the session.

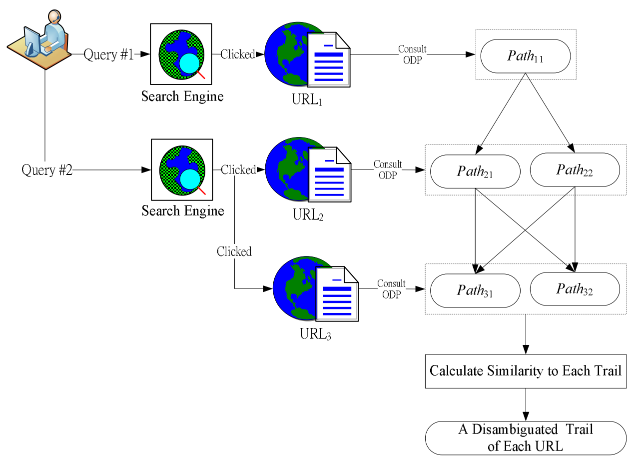

Figure 2 Category Disambiguation Procedure

Table 2. Features for Advertisements Click Prediction

Feature	Description	Feature	Description
Current Impression Level			
QP	Position of q_i in s, i.e., i	QC	ODP categories of query in q_i
#QT	Number of query terms in q_i	QIntent	Intent type of query in q_i
QT	Query terms in q_i	QDMA	DMA level user location ID of q_i
IsURLQ	1 if the query in q_i is in the form of a URL, and 0 otherwise	Qtype	Type of query in q_i: information, navigation, or transaction
QSim	Cosine similarity between query terms in q_i and $q_{i\text{-}1}$	Qoverlap	Overlapping between query terms in q_i and $q_{i\text{-}1}$
First Impression Level			
FQ	Query terms in q_1	TimeToFQ	Time duration (in seconds) between q_1 and q_i
Previous *n* Impression Level (*n*=1,2)			
PNP$_n$	Page number of the result page of $q_{(i\text{-}n)}$	TimeToP$_n$	Time duration (in seconds) between $q_{(i\text{-}n)}$ and q_i
#AdP$_n$	Number of advertisements displayed in the result page of $q_{(i\text{-}n)}$	ClickRP$_n$	Clicked advertisements in the result page of $q_{(i\text{-}n)}$ at each rank
IsClickP$_n$	1 if there is at least one advertisement click in $q_{(i\text{-}n)}$, and 0 otherwise	ClickDNP$_n$	URLdomain names of clicked advertisements in the result page of $q_{(i\text{-}n)}$
AdCP$_n$	ODP categories of the clicked advertisements in $q_{(i\text{-}n)}$	T#ClickP$_n$	Total number of clicked advertisements in $q_{(i\text{-}n)}$
AdIntentP$_n$	Intent types of the clicked advertisements in $q_{(i\text{-}n)}$	#Adoverlap	Displayed advertisements overlapping between $q_{i\text{-}n}$ and $q_{i\text{-}(n\text{-}1)}$
Contextual Impression Level			
T#Ad	Total advertisements reported in $q_1, q_2, ..., q_{(i\text{-}1)}$	CTQIntent	Intent types of queries in $q_1, q_2, ..., q_{(i\text{-}1)}$
T#Click	Total number of clicked advertisements in $q_1, q_2, ..., q_{(i\text{-}1)}$	CTAdC	ODP categories of clicked advertisements in $q_1, q_2, ..., q_{(i\text{-}1)}$
T#ConClick	Total number of advertisements clicked continuously in $q_1, q_2, ..., q_{(i\text{-}1)}$	CTAdIntent	Intent types of clicked advertisements in $q_1, q_2, ..., q_{(i\text{-}1)}$
ConClick	i-j where $q_j, q_{(j+1)}, ..., q_{(i\text{-}1)}$ contain clicked advertisements continuously	CTIntent$_{\text{Dis}}$	Intents of clicked advertisements in $q_1, q_2, ..., q_{(i\text{-}1)}$ after disambiguation
NearClick	i-j where q_j is the nearest impression containing clicked advertisements	CTR	Advertisements click through ratio before q_i = total clicked ads divided by total ads before q_i
CTQC	ODP categories of queries in $q_1, q_2, ..., q_{(i\text{-}1)}$	CTR@m	Click through ratio for each rank in $q_1, q_2, ..., q_{(i\text{-}1)}$
T#Ad@m	Total number of advertisement reports at rank m of $q_1, q_2, ..., q_{(i\text{-}1)}$, where m=1, 2, ..., 8	T#Click@m	Total number of advertisements clicks in each rank of $q_1, q_2, ..., q_{(i\text{-}1)}$

5. EXPERIMENTS AND DISCUSSIONS

In a session, a user submits a sequence of queries intertwined with advertisement clicks. We propose models to predict whether at least one advertisement is clicked before target advertisements are displayed. This problem is defined formally as follows. Given a sequence of impressions ($q_1, q_2, ..., q_{(i\text{-}1)}$), we will predict if there is an advertisement click event when the query in q_i is submitted.

We formulate the predictive problem of advertisement clicks as a supervised learning problem. The click prediction models are constructed by integrating features extracted from the sessions which contain at least one advertisement click on the first 56 days in the AdLogs with four machine learning algorithms including conditional random fields (CRF), support vector machines (SVM), decision tree (DT) and back-propagation neural networks (BPN). All sessions in the last 28 days are used for testing and each impression in a testing session is classified into *click* or *non-click*.

In SVM model, we use RBF and linear kernels. The parameter c of both kernels and parameter *gamma* of RBF are determined by grid algorithm. For the BPN model, several options of the neural network configurations are tested, and the hidden layer, learning rate and momentum are set to 1, 0.8 and 0.2, respectively. For decision tree, we choose the default settings of C4.5. Table 3 shows *Accuracy* (*Acc*), *Precision* (*Prec*), *Recall* (*Rec*) and *F-measure* (*F*1) of non-click (click) prediction. Guessing the majority class (non-click) is one baseline. Markov Model (MM), formulated by query transition, is another baseline.

Table 3. Performance of Advertisements Click Prediction

All Features		Non-click type			Click type		
Model	*Acc*	*Prec*	*Rec*	*F*1	*Prec*	*Rec*	*F*1
Guess	0.9559	0.9559	1.0000	0.9780	0	0	0
MM	0.6917	0.9586	0.7081	0.8334	0.0505	0.3369	0.1937
CRF	**0.8469**	**0.9798**	**0.8575**	**0.9186**	**0.1663**	**0.6167**	**0.3915**
DT	0.8706	0.9666	0.8955	0.9311	0.1270	0.3296	0.2283
BPN	0.8750	0.9672	0.8998	0.9335	0.1344	0.3375	0.2359
SVM (RBF)	0.8809	0.9679	0.9054	0.9366	0.1451	0.3481	0.2466
SVM(Linear)	0.8781	0.9675	0.9028	0.9351	0.1399	0.3431	0.2415

Precision of click prediction is the most important. The CRF model achieving precision of click prediction 0.1663 performs the best. It is better than SVM, DT, BPN and the two baselines significantly (p-value<0.001 in t-test).

In addition, we also compare the performance of some features proposed by Guo *et al*. [11] with our features. Guo *et al*. extracted groups of features such as Query, SERP Content, Result Quality, Interaction, Click and Context from a library search engine. These features are used to predict "Receptive" (i.e., an advertisement click is expected in a future search within the current session) and "Not receptive" (i.e., not any future advertisement clicks are expected within the current session). In their original experiments, four models including CRF (Query), CRF (Query+Click), CRF (All) and CRF (All-Interaction) are reported. The CRF (All) model integrated CRF with all groups of features achieves the best performance such as precision 0.15, recall 0.21 and F-measure 0.17 in the click prediction task. Using the Query and the Click groups of features with CRF performs very close to the best model, especially precision of click prediction (i.e., 0.14). Not all groups of features can be extracted from AdLogs, for example, the logs did not record information about the mouse trajectory. We implement similar prediction models using the two groups of features (Query and Click) extracted from AdLogs for comparison. The results are reported in Table 4. The performance of our models is better than that of Guo *et al*.'s model significantly (p-value<0.001 in t-test) except the recall of click prediction.

Table 4. Performance of Different Models

Comparison		Non-click type			Click type		
Model	*Acc*	*Prec*	*Rec*	*F*1	*Prec*	*Rec*	*F*1
CRF (Our)	0.8469	0.9798	0.8575	0.9186	0.1663	0.6167	0.3915
CRF (Guo)	0.7911	0.9789	0.7987	0.8888	0.1254	0.6261	0.3757
SVM (Our)	0.8809	0.9679	0.9054	0.9366	0.1451	0.3481	0.2466
SVM (Guo)	0.8465	0.9676	0.8685	0.9180	0.1147	0.3693	0.2420

Three feature selection methods, i.e., random subspace (RS) [12], F-score (FS) and information gain (IG), are employed to search for a predictive subset of features. RS is an ensemble model that consists of several classifiers and then prediction is through a majority vote from the classifiers. FS and IG measure the weight of each feature for discrimination. The features of the larger FS and IG weights tend to be more discriminative. Table 5 shows the results. CRF and SVM with RS are better than those with FS, IG or ALL (all features). It shows that classifiers generated by different features and then combined by a majority vote can prevent bias. Integrating 35 classifiers of the CRF model achieves accuracy 0.8511 and precision of click prediction 0.1721. The performance of the best model is better than that of the other models significantly (p-value<0.001 in t-test)

Table 6 shows the relative importance (RI) and feature level (FL) of the top-10 features for FS and IG. The weight of relative importance for each feature is normalized by the highest weight. FL denotes the feature level from which features are selected. Here, CI, FI, PI, and CT denote Current Impression Level, First Impression Level, Previous *n* Impression Level, and Contextual Impression Level, respectively. Features belong to CT is very import because more than half of the top-10 important features are selected from this level. RS is neglected in the discussion due to its feature sets are random sampled.

Query term (QT) is the most important feature in both feature selection algorithms and the first query (FQ) in a session is selected by F-Score. We find that advertisements are usually clicked after some specific queries such as "eBay" and "Walmart". However, if a prediction model only uses query terms, the performance is not good. MM demonstrated such a case. Notably, the prediction model needs to consider more predictive features to improve the prediction performance.

User click behaviors such as CTR, IsClickP_1, ConClick, and NearClick are good predictors. According to click probability, if a user is clicking on an advertisement, s/he is more likely to click on other advertisements in the future.

Some intent-related features such as CTAdIntent, CTQIntent and Qtype are selected by both feature selection algorithms. It may be due to the fact that some sort of users' intent is useful for predicting advertisement clicks. In line with previous study [2], we find that the click through ratio of navigational query is higher than informational query. The features derived from the intent clusters of query and clicked advertisement rank very prominently. We observe that intent types identified as "shopping" and "recreation" have more advertisement clicks than other types. User intent disambiguation is important as well because more clearly intent will be identified after the disambiguation procedure.

Table 5. Performance of Feature Selection

Features Selection		Non-click type			Click type		
Model	*Acc*	*Prec*	*Rec*	*F*1	*Prec*	*Rec*	*F*1
CRF(ALL)	0.8469	0.9798	0.8575	0.9186	0.1663	0.6167	0.3915
CRF(RS15)	0.8457	0.9797	0.8563	0.9180	0.1648	0.6145	0.3897
CRF(RS25)	0.8493	0.9801	0.8598	0.9199	0.1696	0.6210	0.3953
CRF(RS35)	**0.8511**	**0.9803**	**0.8615**	**0.9209**	**0.1721**	**0.6242**	**0.3982**
CRF(RS45)	0.8504	0.9802	0.8609	0.9205	0.1711	0.6230	0.3971
CRF(FS)	0.8473	0.9799	0.8579	0.9189	0.1670	0.6175	0.3923
CRF(IG)	0.8479	0.9799	0.8585	0.9192	0.1678	0.6186	0.3932
SVM(ALL)	0.8809	0.9679	0.9054	0.9366	0.1451	0.3481	0.2466
SVM(RS15)	0.8796	0.9677	0.9042	0.9359	0.1426	0.3457	0.2442
SVM(RS25)	0.8811	0.9679	0.9057	0.9368	0.1456	0.3486	0.2471
SVM(RS35)	0.8813	0.9679	0.9058	0.9369	0.1459	0.3488	0.2474
SVM(RS45)	0.8815	0.9679	0.9060	0.9370	0.1463	0.3492	0.2477
SVM(FS)	0.8811	0.9679	0.9056	0.9368	0.1455	0.3485	0.2470
SVM(IG)	0.8812	0.9679	0.9058	0.9368	0.1458	0.3488	0.2473

Table 6. Top-10 Important Features

	F-Score			Information Gain		
Rank	Feature	FL	RI	Feature	FL	RI
1	QT	CI	1	QT	CI	1
2	CTAdIntent	CT	0.7751	$CTIntent_{Dis}$	CT	0.6284
3	$CTIntent_{Dis}$	CT	0.6498	CTQIntent	CT	0.5268
4	CTQIntent	CT	0.5092	T#ClickP_1	PI	0.4128
5	FQ	FI	0.3557	CTR	CT	0.2884
6	IsClickP_1	PI	0.3222	T#Ad	CT	0.2612
7	CTR	CT	0.3052	ConClick	CT	0.2475
8	T#ClickP_1	PI	0.2943	CTAdIntent	CT	0.2386
9	ConClick	CT	0.2688	NearClick	CT	0.2179
10	NearClick	CT	0.2568	Qtype	CI	0.2082

We summarize our observations as follows. Query terms are important, but should be used along with other features. Users' click behaviors are good predictors for improving predictive performance. Most of the selected important features are related to user intent. That demonstrate the intent-related features are very powerful for advertisement clicks prediction. Larger number of important features comes from contextual impression level. That shows advertisement clicks behavior is not discrete activities in a session and contextual information is helpful for identifying clear user intent for better advertisement click prediction.

6. CONCLUSIONS AND FUTURE WORK

This paper learns users' behaviors from search and click logs. We explore the effects of various intent-related features on advertisements click prediction. CRF model performs better than two baselines and SVM significantly. When random subspace method is introduced to feature selection, the precision of click prediction is increased from 0.1663 to 0.1721. We also investigate important features in the click prediction task and most of them are from contextual impression level. That demonstrate the user contextual behavior is very useful for prediction. In the future, we plan to expand our model to consider fine-grained user intent and user interactions. In addition, we will extend this approach to predict which advertisements will be clicked.

7. ACKNOWLEDGMENTS

This research was partially supported by National Science Council, Taiwan under grant NSC99-2221-E-002-167-MY3. We are also grateful to Microsoft Research Asia for the support of MSN Search Query Log excerpt.

8. REFERENCES

[1] Abhishek, V. and Hosanagar, K. 2007. Keyword generation for search engine advertising using semantic similarity between terms. In *Proceedings of the 9th international conference on Electronic commerce*, 89 – 94.

[2] Ashkan, A., Clarke, C.L., Agichtein, E. and Guo, Q. 2009. Classifying and Characterizing Query Intent. In *Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval*, 578 – 586.

[3] Benjamin E., Michael O. and Michael S. 2007. Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords. *American Economic Review*. 97, 1, 242-259.

[4] Boldi, P., Bonchi, F., Castillo, C., Donato, D., Gionis, A. and Vigna, S. 2008. The query-flow graph: model and applications. In *Proceeding of the 17th ACM conference on Information and knowledge management*, 609-618.

[5] Broder, A. 2002. A taxonomy of web search. *SIGIR Forum*. 36, 2 (2002), 3-10.

[6] Broder, A., Ciaramita, M., Fontoura, M., Gabrilovich, E., Josifovski, V., Metzler, D., Murdock, V. and Plachouras, V. 2008. To swing or not to swing: learning when (not) to advertise. In *Proceeding of the 17th ACM conference on Information and knowledge management*, 1003 – 1012.

[7] Buehrer, G., Stokes, J.W. and Chellapilla, K. 2008. A large-scale study of automated web search traffic. In *Proceedings of the 4th international workshop on Adversarial information retrieval on the web*, 1 – 8.

[8] Cheng, H. and Cantú-Paz, E. 2010. Personalized click prediction in sponsored search. In *Proceedings of the3rd ACM international conference on Web search and data mining*, 351 – 360.

[9] Craswell, N., Zoeter, O., Taylor, M. and Ramsey, B. 2008. An experimental comparison of click position-bias models. In *Proceedings of the international conference on Web search and web data mining*, 87 – 94.

[10] Guo, Q. and Agichtein, E. 2010. Exploring searcher interactions for distinguishing types of commercial intent. In *Proceedings of the 19th international conference on World wide web*, 1107 – 1108.

[11] Guo, Q. and Agichtein, E. 2010. Ready to buy or just browsing?: detecting web searcher goals from interaction data. In *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, 130 – 137.

[12] Ho, T.K. 1998. The Random Subspace Method for Constructing Decision Forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 20, 832 – 844.

[13] Jansen, B.J., Booth, D.L. and Spink, A. 2007. Determining the user intent of web search engine queries. In *Proceedings of the 16th international conference on World Wide Web*, 1149-1150.

[14] Jelinek, F. and Mercer, R. 1980. Interpolated estimation of Markov source parameters from sparse data. In *Proceedings of the Workshop on Pattern Recognition in Practice*.

[15] Kohavi, R. and John, G.H. 1997. Wrappers for feature subset selection. *Artificial Intelligence*. 97, 273 – 324.

[16] Lacerda, A.,Cristo, M., Gonçalves, M.A., Fan, W., Ziviani, N. and Ribeiro-Neto, B. 2006. Learning to advertise. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, 549 – 556.

[17] Liu, H. and Motoda, H. 1998. *Feature Selection for Knowledge Discovery and Data Mining*.

[18] Manshadi, M. and Li, X. 2009. Semantic tagging of web search queries. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural* Language *Processing of the AFNLP: Volume 2 - Volume 2*, 861-869.

[19] Nguyen, V. and Kan, M. 2007. Functional Faceted Web Query Analysis. *Query Log Analysis: Social And Technological Challenges. A workshop at the 16th International World Wide Web Conference*.

[20] Perugini, S. 2008. Symbolic links in the Open Directory Project. *Information Processing and Management: an International Journal*. 44, 2 (2008), 910-930.

[21] Radlinski, F. and Joachims, T. 2005. Query chains: learning to rank from implicit feedback. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, 239 – 248.

[22] Richardson, M., Dominowska, E. and Ragno, R. 2007. Predicting clicks: estimating the click-through rate for new ads. In *Proceedings of the 16th international conference on World Wide Web*, 521 – 530.

[23] Schlosser, A.E., Shavitt, S. and Kanfer, A. 1999. Survey of Internet *users'* attitudes toward Internet advertising. *Journal of Interactive Marketing*. 13, 3, 34-54.

[24] Wang, C.J., Lin, K. and Chen, H.H. 2010. Intent boundary detection in search query logs. In *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, 749-750.