What Are Driving Users to Click Ads? User Habit, Attitude, and Commercial Intention

Jun Yan
Microsoft

Ning Liu
Nanjing Medical University

Weiguo Fan
Virginia Tech

What Are Driving Users to Click Ads? User Habit, Attitude, and Commercial Intention

Siyu Gu[1*] Jun Yan[2] Zhanyong Xiao[1] Ning Liu[2] Dou Shen[3]
Weiguo Fan[4] Lei Ji[2] Zheng Chen[2] Ying Chen[1]

[1]Beijing Inst. of Tech
No5, ZhongGuanCun NanDaJie
Beijing, P. R .China 100081
{gusuperstar, xiaozhanyong, chengying1}@bit.edu.cn

[2]Microsoft Research Asia
Sigma Center, 49 Zhichun Road
Beijing, P.R. China 100190
{junyan, ningl, leiji, zhengc}@ microsoft.com

[3]Virginia Tech
Dept. of Computer Science
Blacksburg VA, USA
wfan@vt.edu

[4]Microsoft Corporation
One Microsoft Way,
Redmond WA 98052, USA
doushen@microsoft.com

ABSTRACT
With the rapid growth of World Wide Web (WWW), online advertising is attracting more and more attention owing to its huge market potentials. In fact, in general sponsored search, the problem that what drive users to click ads is still underexplored in Web data mining, which essentially prevents us from further exploiting this market. In this paper, we propose to learn the factors that drive users to click ads by a large scale data driven study, on the basis of the insights researched by psychological studies which have been taken to understand user intentions. Three psychological factors are reformulated and verified in this work. They are *user habit*, *attitude* and *commercial intention*. We first introduce how to extract features according the approximate effect of these three psychological factors from the click-through log data of a commercial search engine. Then, we analyze the correlation between users' ad click behaviors and those factors. After that, as an application, we build a prediction model integrating these factors to predict the probability that whether the user will click ads. The result of experiments on real user ad click log of a commercial search engine shows the prediction model can help classical machine learning algorithms perform better.

Categories and Subject Descriptors
H.3.5 [**Information Storage and Retrieval**]: Online Information Service – *commercial service*; I.6.4 [Computing Methodologies]: SIMULATION AND MODELING – Model Validation and Analysis;

Keywords
Online advertising, sponsored search, user habit, user attitude, user commercial intention

1. INTRODUCTION
Nowadays, a large number of Internet companies are investing significant amount of money for online advertising, as the Internet has become the mainstream in daily lives. For search engines, the economic impact of sponsored search is immense, while user's ad clicks are the main source of revenue. Numerous studies have focused on how to promote revenue by optimizing the relationship between ad keyword bid price and ad relevance [17, 24, 39]. However, the problem of what are driving users to click ads is still underexplored, which greatly prevents us from further developing this online advertising market.

To understand users' behaviors such as using Web services and clicking ads, there are a large number of research works in psychological studies that have attempted to predict and explain human behaviors [1, 6, 33]. The Theory of Reasoned Action (TRA) [18] and Theory of Planned Behavior (TPB) [2] are two of the well known models in the field of psychology. However, all these studies are accomplished through questionnaires, which are not of large scale and hard to be generalized to real world Web applications. To our best knowledge, the work by Liu *et al.* [26] is one of the earliest data driven studies that use search toolbar data to estimate the effectiveness of sponsored search advertising. Though their observations learnt from toolbar data have proved that users are prone to accept ads, few of the previous data driven studies have analyzed the factors that affect users' online ad click behaviors. To fill the gap between psychological studies and Web data mining studies on the problem of what are driving users to click ads, in this paper, we propose to reformulate and verify three psychological factors that may affect users' ad clicks.

The three psychology factors we propose to study are *user habit, attitude* and *commercial intention*. In the psychological definition, user habit is a kind of action tendency, which is learned from the historical user behaviors. For example, if a user always tends to click ads under some conditions without subjective ad click intention, we say this user has formed an ad click habit under this condition. In terms of Web data mining, user habit is a factor similar to the prior probability of user click ads. However, it differs from the prior probability since it can estimate the stability of the user behavioral tendency. Attitude is defined as the degree of favor or disfavor of a user to a particular entity. In ad click analysis, a user's attitude towards ads is determined by the ad relevance pertaining to the user's query, the positive feeling to certain terms in ad title etc. The commercial intention is defined as the perceived business purpose in psychology [36]. In Web search, commercial intention reflects whether a user has intention to purchase or participate in commercial services [14]. In this work, we propose a set of data driven features, which can be learned from the search engine logs, to approximate these three psychological factors. Then, we show the strong correlation between users' ad click behaviors and these three factors through

ADKDD'10, July 25, 2010, Washington D.C., USA.
Copyright 2010 ACM 978-1-4503-0221-0…$10.00.

[1]* This work was done when the author was visiting Microsoft Research Asia.

empirical studies over the real ad click-through log from a commercial search engine.

Besides independently analyzing these three factors, which may drive users to click ads, we build an ad click prediction model to integrate the information coming from these three factors. Motivated by the basic assumption of classical click models [8, 13, 15], *there is a click only if the user examines the url and was attracted by it*, in our model, habit and commercial intention are assumed to determine whether the user will examine the displayed ads, and the attitude of users towards ads determines whether the user will click the ads. In this work, we do not aim to state that users' habit, attitude and commercial intention are the only factors that affect user's click behaviors. We aim at providing sources of evidences to show their correlations to users' ad click behaviors and show their power in predicting users' ad click behaviors. Thus as an application, we use these factors to predict whether the user will click ads within a specific page view. The results show that we can better predict whether a given user will click ads by combing these three factors through our proposed prediction model, compared with classical features and machine learning algorithms for this problem.

In the rest of this paper, Section 2 defines and formulates the user habit, attitude and commercial intention. Section 3 proposes an ad click prediction model by integrating three factors studied. Section 4 introduces the experimental results on predicting. Section 5 summarizes some related works about ad click prediction and psychological user models. Section 6 concludes this work and provides some future research lines.

2. Psychological factors

To introduce these three psychological factors in a data driven manner, we take the ad click-through log of a commercial search engine as data source. The log data contains the search queries, the displayed ads, and the clicked ads of each user session. The detailed data format is summarized in Table 1. In our study, we use the click-through log data ranging from June 1st to August 31st, 2009. There are 10,000 randomly sampled users and 250,451 unique ads in the dataset. Because we want to learn users' habits, we divide this data set into two sub-datasets: history dataset and future dataset. History dataset is used for habit learning while future dataset is used for prediction. We set the data in June and July as "history", while the data in August as "future".

Table 1. Click-through log data format used in our study

Name	Explanation
UserID	One unique user ID for each unique user.
QueryText	The text content of the query used by the user.
QueryTime	The time when the query was issued.
DisplayAd	The URL list of all the ads displayed by the query in the same ad display page.
ClickAd	The ad URLs clicked by the user.
ClickTime	The time when the click event occurred after the query was issued

Based on this data set, we reformulate and verify these three factors, in the following three sub-sections respectively to show their correlations with users' ad click behaviors.

2.1 User Habit

Judith et al. [27] defined user habit as the customary ways of behaving, and the relatively bare-bone definition of habit is behavior tendencies. They are tendencies to repeat responses given a stable supporting context. With repetition and practice, the cognitive processing becomes automatic and can be performed quickly with allocation of minimal focal attention. Moez et al. [25] also defined habit as automatic behavior tendencies developed during the past history of the individual. Inspired by the psychological definitions on it, in this section, we propose a set of novel data driven features, which could describe the users' ad click tendency based on their historical ad click behaviors.

For each user u, we use a vector $C_u = (c_{u1} \ldots c_{uT})$ to represent the user's ad click history, where T is the length of history, i.e. number of time stamps we used for user habit learning. c_{ui} stands for the number of ad clicks conducted by user u at timestamp i. A time window W is defined as a sequence of continuous time stamps with fixed length, which is designed to control the length of time segment. The user habit features are defined below.

Wratio: (Window Ratio) given a window size S, a valid time window is defined as the time windows within which the user clicked at least one ad We denote a time window W of user u starting from time stamp i with length S by the number of ad clicks conducted by u as,

$$W_i(u) = \{c_{ui} \ldots c_{u(i+S-1)}\}.$$

Then the valid window is defined as,

$$Valid(W_i(u)) = \begin{cases} 0 & \forall t, c_{ut} = 0, i \le t \le i+S-1 \\ 1 & \exists t, c_{ut} > 0, i \le t \le i+S-1 \end{cases} \quad (1)$$

Wratio is defined as a feature describing the ratio of valid time windows over all the time windows considered.

$$Wratio = \frac{\sum_{i=1}^{i=T-S+1} Valid(W_i)}{T-S+1} \quad (2)$$

If most of the time windows are valid time windows, it suggests a strong ad click habit of the user, because for every S day, the user will repeat the action of ad click at least once.

Entropy: In information theory, entropy is a measurement for the uncertainty associated with a random variable. Based on our symbols, the definition of ad click Entropy is:

$$p_i = \frac{\sum_{j=(i-1)*S+1}^{j=i*S} sign(c_{uj})}{\sum_{i=1}^{i=T/S} \sum_{j=(i-1)*S+1}^{j=i*S} sign(c_{uj})}$$

$$Entropy = -\sum_{i=1}^{i=T/S} p_i \log(p_i) \quad (3)$$

where $$sign(c_{uj}) = \begin{cases} 1 & c_{uj} > 0 \\ 0 & c_{uj} = 0 \end{cases}$$

This feature estimates the distribution of ad click in users' historical online ad click behaviors. The more uniform the user's historical ad click distribution is, the more stable the user's ad click habit is.

Enfor: This is another intuitive understanding of user habit. If the same user performs ad click events continuously, the strength of the ad click behavior tendency would be stronger. Contrarily, if a user did not repeat a behavior for a long time, we are less confident to say that this user has the habit of click ads. Thus, Enfor describe that whether the user is accustomed to click ads since it analyzes user history. To estimate the different effects of different time periods in history, we divide the historical time periods into two kinds of segments, which are positive segments and negative segments. The positive segment is defined as the largest continuous time stamps containing ad clicks, and the negative segment is defined as the largest continuous time stamps not containing ad clicks. For example, if the user's ad click history is (c_{u1}=1, c_{u2}=1, c_{u3}=0, c_{u4}=0, c_{u5}=0, c_{u6}=1, c_{u7}=1, c_{u8}=1), the positive segments are ($c_{u1}c_{u2}$), ($c_{u6}c_{u7}c_{u8}$) and the negative segments are ($c_{u3}c_{u4}c_{u5}$). We denote the set of all segments as *SEG* = $\{s_1, s_2, \ldots s_k\}$, where s_i is either a positive segment or a negative segment. A longer positive segment will lead to a larger contribution for habit building, and a longer negative segment will lead to a larger effect of forgetting. We assume the rate for habit enhancing and forgetting is generally stable but according to a different factor. Given the enhancing factor $EN \in R$ and forgetting factor $FOR \in R$, we define Enfor feature as,

$$Positive(s_i) = \begin{cases} 1 & s_i \, is \, a \, positive \, segment \\ 0 & otherwise \end{cases}$$

$$Negative(s_i) = \begin{cases} 1 & s_i \, is \, a \, negative \, segment \\ 0 & otherwise \end{cases}$$

$$Enfor = \sum_{i=1}^{i=k} Positive(s_i)|s_i|^{EN} - Negative(s_i)|s_i|^{FOR} \quad (4)$$

2.2 Attitude

In psychology, attitude is a psychological tendency expressed by evaluating a particular entity with some degree of favor or disfavor [16]. A utilitarian definition of attitude is a predisposition to behave in a certain way [3]. It guides behavior in ways that are beneficial to the individual [32]. In our problem, we assume that whether the user will click ads will be determined by the attitude of the user to the displayed ads. For example, relevance of ads is one determinant of attitude since more relevant ads could better satisfy user's needs and then lead to more positive attitude. There are also some other determinants to affect the user's attitude, such as the attractiveness of the ads and the popularity of the ads' domain. Motivated by this, we formulate a set of features for describing user attitudes.

In the work taken by Matthew *et al.* [31], a number of features have been designed based on characteristics of ads, which can be classified into three categories: title features, URL features and relevance between ads and queries. All the three categories of features could be considered as attitude features since both ad attractiveness and ad relevance may significantly affect users' attitude to the ads. In other words, we consider most of classical ads related features in previous works [31] as the attitude features. In this work, we propose to reformulate some classical ones in each category for describing users' attitude to ads.

For the ad title based features, we assume that the terms in the ad title determine the attractiveness and the appearance of ads, which will be the key factors to affect users' attitude toward the ads. Matthew *et al.* [31] implemented a unigram feature for each of the most common 10,000 words in ad title and ad body in a collection. In this way, there are more than ten thousands sparse features. To reduce the number of features and retain most of the characteristics, we propose to learn the ad CTR of each term in ad title from ad click through log. The CTR is defined as follows

$$CTR(e) = \frac{\#clicks \, of \, all \, ads \, whose \, title \, contains \, e}{\# \, displays \, of \, all \, ads \, whose \, title \, contains \, e} \quad (6)$$

where *e* is a term in an ad title. The higher *CTR(e)* means it is helpful for the ad to attract ad click when the term *e* appears in the ad title. Some high CTR terms are "site:", "famous", "good", "more", "trade", "marketplace", "old", "cloth". etc. We reformulate several classical features for describing the ad attractiveness, which determine the user attitude in turn. The features we used in this study are listed in Table 2.

Table 2. Ad title based features

Feature	Description
T_total	The sum of term CTRs of all the title terms in the specific ad display session.
T_ave	The average CTR of all the title terms in the specific ad display session.
T_max	The maximum CTR among all the title terms in the specific ad display session.

Similarly, we propose to learn the CTR of each ad domain, which is defined as the domain of the ad URL. Some domains are more popular to users, so there is a higher probability that users have better attitudes toward these ads and then click the ads. For example, if an ad comes from "eBay", users tend to have good attitude to it owing to the good reputation of the eBay Web site. The extracted features in this work are listed in Table 3.

Table 3. Ad domain based features

Feature	Description
D_total	The sum of domain CTRs of all the ad domains in the specific ad display session.
D_ave	The average CTR of all ad domains in the specific ad display session.
D_max	The maximum CTR among all the ad domains in the specific ad display session.

Ad relevance [29], which is known as the content relevance between a query and an ad, has been considered as the key factor that affects ad click-through rate. From our point of view, it is also a key factor to affect users' attitudes since we assume the users will prefer to click the relevant ads to their current intent, i.e. their current query. To estimate the relevance between ads and queries, we leverage some classical ad relevance features [31], which are summarized in Table 4. In the table, CQR is defined as the number of query terms contained in the ad's title divided by the number of terms in the query, and CS is defined as the cosine similarity between the query and the ad.

Table 4. Ad relevance features

Feature	Description
Connum	The ratio of ads whose title contains at least one term in user query.
Callnum	The ratio of ads whose title contains all of the terms in query.
Conave	The average CQR among all the ads in a specific ad display session.
Maxcon	The maximum CQR among all the ads in a specific ad display session.
f_rel	The average CS among all the ads in a specific ad display session.
f_mrel	The maximum CS among all the ads in a specific ad display session.

2.3 Commercial Intentions

Another important factor, which can drive users' ad click, is Online Commercial Intention (OCI). It reflects whether a user has intention to purchase or participate in commercial services [14]. Our assumption is that commercial intention drives the user to examine and click ads more possibly. We introduce the features that are used to represent users' commercial intentions in next sub-section.

In sponsored search, query is the main trigger for selecting ads to display. In this work, the commercial intents impacting on ad click are leveraged from the users' queries. Dai *et al.* [14] claimed some queries can be detected with their commercial intents directly since they have explicit indicators in the queries, such as "price", "cheap", "buy", "sell", etc. In our investigation, we also use these explicit indicators to represent commercial intents of users. If the query contains one of the indicators, we consider when user searches the query, he/she has commercial intention. The top 20 indicator terms is shown in Table 5.

Table 5. Top 20 indicator terms for commercial intention

scottrade	***sale***	***coupon***	***hire***
hsn	***overstock***	***rental***	***price***
JCPenney	***cheap***	***fare***	***rebate***
ebay	***lookup***	***size***	***depot***
ticket	***discount***	***retail***	***dollar***

To verify whether commercial intents inspire ad click, we compared the CTR on the queries with commercial intents and that of the general queries. The Commercial CTR and General CTR are defined as follows:

$$Commercial\ CTR = \frac{\#ad\ clicks\ \ by\ commercial\ query}{\#ad\ displays\ by\ commercial\ query} \quad (7)$$

$$General\ CTR = \frac{\#ad\ clicks\ by\ any\ query}{\#ad\ displays\ by\ any\ query} \quad (8)$$

In our dataset, there are 1,210,00 pages. We perform 10 folds cross validation. The results of CTRs are in Table 6. From this table, we can see that averagely the Commercial CTRs are larger than the corresponding General CTRs, which means commercial intentions inspire ad click significantly. In addition, the *t*-test result $p<0.001$ shows that the difference is statistically significant.

Table 6. Compare CTR between the Commercial and General queries.

	Average	Standard dev
Commercial CTR	0.027094	0.002629
General CTR	0.022244	0.000994

Table 7. Commercial intention features

Feature	Description
Q_total	The sum of commercial scores of all the terms in the query.
Q_ave	The average commercial score of every term in the query.
Q_max	The maximum commercial score of terms in the query.

For feature extraction, we set the history term CTR as the commercial score of each term. Based on Table 5, we extract three features to estimate the user's commercial intentions. These three features are only for demonstration purpose since there will be much more indicator terms and more features that can be used to represent users' commercial intents. The three features are described in Table 7.

3. AD CLICK PREDICTION MODEL

To show the between of factors and users' ads click, our problem is configured as to predict whether a user will click an ad in one ad impression, which is different from classical ad click models. In previous works, many of them aim to solve the position bias among the clicks within one session [8, 13, 15]. Experiments in our study show that user habit, attitude and commercial intention have strong correlations with ad display session CTR. Thus we propose to build a model that can combine these three factors to predict whether a user will have ad click in one impression.

One of the basic assumptions of classical click models is that a click to an object happens only if the user examined the object to be clicked and was attracted by it [8, 13, 15]. In our proposed model, our basic assumption is that whether the user will examine the ads displayed in the same page is determined by the user's habit to ad click and commercial intention, while the user's attitude toward the ads determines whether she will click one of them after examining them. As an example for intuition, if a user has the habit to examine the ad area in sponsored search or has the commercial intent to view some ads in current session, then she/he has high probability to examine the ads. After examining any of the ads, if she/he finds that the ads have high probabilities to satisfy her/his intention, she/he will have positive attitude and will have high probability to click one of the ads. The work in [21] showed that the extent to check sponsored ads is different among users. This is pertaining to user's habit, which is reflected in the user's historical behaviors. Users do not always click ads after they examine them. A click happens if and only if the user's attitude towards the ad is positive after examining it. Based on these assumptions, the ad click prediction model is shown in Figure 1. The basic assumptions of this Bayesian network are described by the equations below:

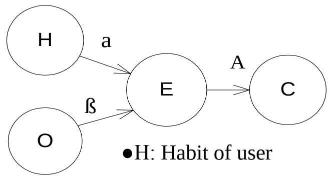
$$P(E=1) = H*\alpha + O*\beta \quad (9a)$$

$$P(C=1|E=1) = A \quad (9b)$$

$$P(C=1) = P(C=1|E=1)P(E=1) \quad (9c)$$

$$P(C=1) = (H*\alpha + O*\beta)*A$$
$$= \alpha*(H*A) + \beta*(O*A) \quad (9d)$$

In the combined model, α estimates the probability to examine ads if the user had habit and β estimates the probability to examine ads if the user had commercial intention. In a specific user session, the probability of whether the user will examine ads is estimated by (9a). After the user examines any of the ads, the general attitude towards the ads determines whether she/he wants to click any of the ads (9b). Although the user may have different attitude to different ad, we combine them into one general score, because in this study we do not care about which specific ad the user clicks. Thus the probability whether the user will click ads in a specific ad display session is estimated by (9c). After substituting the factors in (9c), we obtain a final formula (9d).



●H: Habit of user

●O: Commercial intention of user

●E: did the user examine the ads?

●C: did the user click any ads?

●A: Attitude of user

Figure 1. The ad click prediction model

For H, O and A, we extract the features from click-through data as described in Section 2. Then we use linear regression to obtain the general scores. The ground truth for model training is the sessions with clicks observed from click-through data. C is binary, denoting whether the user clicked ads in that ad display session. After we get H, O and A, the parameters α and β can also be estimated using linear regression since P(C=1) is the linear combination of *H*A* and *O*A*. After the parameters are estimated, given the inputs H, O and A for each ad display session, the model outputs a score P(C=1), representing the probability for the user to click ads in that session.

4. EVALUATION

In this section, we introduce our experimental results. The habit features are extracted from history data set while the attitude and commercial intention features extracted from future dataset. We introduce the effect of each factor in Section 4.1. And, in Section 4.2, we introduce the results of ad click prediction.

4.1 Factor Estimation

4.1.1 For Ad Prediction

To estimate the effect of each factor toward ad click prediction, we directly use the Multiple Linear Regression (MLR) [4] to learn the combination. Multiple Linear Regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. It estimates the relationship between a dependent variable and one or more independent variables. Since the ad click prediction problem is configured as a binary classifcation problem, we set different thresholds to the regression score to determine whether the output is positive or negative. To estimate the performance of the classification, we use the Receiver Operating Characteristic (ROC) curve [42] analysis as the evaluation metric. ROC curve is to represent the trade-off between the false positive and true positive rates for every possible cut off (classification threshold). Here the true positive rate and the false positive rate are defined as,

$$true\ positive\ rate = \frac{true\ positive}{page\ views\ containng\ ad\ clicks} \quad (10)$$

$$false\ positive\ rate = \frac{false\ positive}{page\ views\ not\ containing\ ad\ clicks} \quad (11)$$

where true positive is the page views containing ad clicks and predicted as positive rightly, while false positive stands for the page views not containing ad clicks, but is wrongly predicted as positive. In this experiment, we aim to compare the ad click prediction results by using each single factor. For each factor, we do a linear regression to learn the weight of each feature for the combined score of each factor. In addition, we directly use the MLR to learn the combinarion of these three factors. The detailed results are shown in Figure 2.

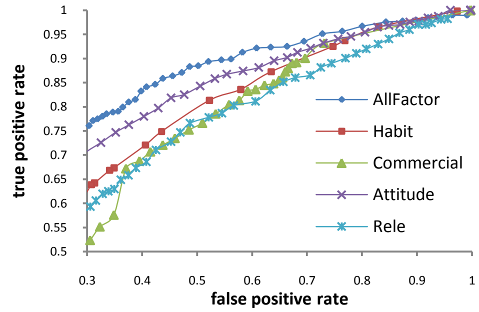


Figure 2. Comparing factors to predict ad click

From Figure 2 we can see that the result using all the three factors "AllFactor" is the best, which means it is helpful for ad click prediction by combing these three factors. Since in the work of [9] ad relevance between the ads and user query has been considered as the key factor that affects ad click-through rate, we also compare the results of considering each factor independently with the ad relevance, which is considered as baseline "Rele". From the figure we can see that besides the combined results, the attitude score and habit score can both significantly outperform the prediction power of ad relevance. The commercial intention is comparable with the ad relevance. As a conclusion, all these three factors have some degree of power in prediction users' ad click, and the combination of them can give outstanding performance compared with ad relevance.

4.1.2 The PLS Path Modeling for Evaluation

In psychology, Partial Least Squares (PLS) is a widely used class of methods for modeling relations between sets of observed variables by means of latent variables. PLS has attracted much attention in recent years [40] since it has the ability to model latent structures [10, 11, 12], and "involves no assumptions about the population or scale of measurement" [19]. A PLS path model is described by two models: (1) a measurement model relating the Manifest Variables (MVs), *i.e.* observable variables, to their own Latent Variables (LV), *i.e.* unobservable variables; and (2) a structural model relating some LVs to other LVs. In this work, it is supposed that the LV ξ is generated by its own MVs. The LV ξ is a linear function of its MVs plus a residual term

$$\xi = \sum_h w_h x_h + \delta \quad (12)$$

where x_h is the MV value, w_h is the weight of x_h, and δ is the associated measurement error. The linear equations that deal with the relationship between LVs are

$$\xi_j = \beta_{j0} + \sum_i \beta_{ji}\xi_i + \boldsymbol{v}_j \quad (13)$$

where ξ_i denotes the LV whose subscript is $\boldsymbol{i}$, β_{ji} denotes the coefficient between ξ_j an ξ_i, and $\boldsymbol{v}_j$ is the random disturbance term.

The estimation of the weights and the structural equations using PLS is introduced in detail in [37]. There are two major metrics in PLS results. The first metric is regression coefficient, the larger the regression coefficient is, the stronger impact between the two latent variables is. To confirm whether the regression coefficient is statistical significant, the cross-validation methods like bootstrap in PLS-Graph for t-statistics [37] can be utilized. The larger the t-statistics is, the better the result is. The second metric is the R^2 which estimates the proportion of variability in the data set accounted for by the model. This provides a measure of how well future outcomes are likely to be predicted by the model. Also, the larger the R^2 is, the better results we have achieved.

In our study, the three psychology factors are all latent variables, which cannot be observed directly. To estimate the effect of each factor for ad click prediction, we use PLS regression treating the features extracted from click-through data as the indicators of the corresponding factors. The model built for estimation uses each factor alone to predict ad click or uses some combination of the factors to predict ad click. As an example, the power of the model using Habit to predict AdClick is shown in Figure 3. The indicator of AdClick is a binary score denoting whether the user clicked ads in the ad display session.

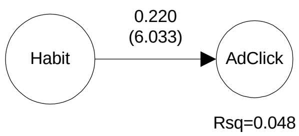


Figure 3. PLS regression for user habit factor

Table 8. PLS regression results

(*:p < 0.1; **: p < 0.01; ***: p< 0.001)

Factor		Regression Coefficient	R^2
Habit		0.220 (6.033)***	0.048
Attitude		0.245 (6.524)***	0.060
Commercial intention		0.199 (4.108)***	0.040
Combined Model	**Habit**	0.180 (3.026)**	0.112
	Attitude	0.199 (4.360)***	
	Commercial Intention	0.125 (1.939)*	
Reduced Model	**Habit**	0.189 (5.189)***	0.092
	Attitude	0.212 (5.921)***	

In Figure 3, the regression coefficient is 0.220, which means habit has a positive effect on ad click. The t-test score of the coefficient is 6.033. The coefficient is significant and it is at the 0.001 level. The R-square is 4.8, which means the percent of the AdClick variance expressed by this model is 4.8. Similarity, we also build the model of using attitude, commercial intention respectively. The PLS regression results are shown in Table 8. In addition, we use a combined model, which considers the combination of all three factors, and a reduced model, which exclude the commercial intention from the model because we observe in the combined model the coefficient between commercial intention and AdClick is not significant. In Table 8, the most left column denotes the model and the corresponding factors. In the regression coefficient column, the first number denotes the coefficient between the factor and AdClick, and the number in the brackets denotes the t-statistics score, which estimates the significance of the coefficient. The last column of the table stands for the R^2 results. The right upper asterisks denote the significance level, while *: p < 0.1; **: p < 0.01; ***: p< 0.001.

In contrast to the results of habit factor, for the attitude model, the path coefficient is 0.245, and the t-test score is 6.524, at the 0.001 level. The result of commercial intention is similar with that of habit. In the combined model, the percentage of AdClick variance expressed by the model increases to 11.2, which is much larger than the model using each factor independently. The coefficient between commercial intention and ad click is at the 0.1 level, and we cannot validate the significance of the coefficient in this case. However, when we only use habit and attitude to predict AdClick, the percentage of variacne accounted by the model decreases to 9.2, and all the coefficients are siginficant. This provides evidence for that the commercial intention also plays an important role in the combined model.

From the PLS results we can see that user habit, attitude and commercial intents all have important impacts on ad click. We can also see that when we use these three factors simultaneaously it has a much stronger capability to predict ad click than using only one factor.

4.2 Unified Model for Ad Click Prediction

In this subsection, we show the effect of our proposed features and unified model on enhancing the result of classic method.

4.2.1 Baseline Algorithms

We compare our model with classical machine learning algorithms, which directly consider all features listed in this paper as inputs. The baseline classification algorithm we used is Multiple Linear Regression.

- **Multiple Linear Regression (MLR)**

The baseline algorithm for our experiment is Multiple Linear Regression, which has been introduced in Section 4.1.1.

4.2.2 Evaluation Metrics

- **ROC Curve**

The ROC curve is the first evaluation metric we will use in this experiment. To guarantee the stability of our experiments, we take cross-validation to verify our model. Cross-validation is a technique for assessing how the results of a statistical analysis will generalize to an independent data set, and it is mainly used for estimating how accurately a predictive model will perform in practice. In *K*-fold cross-validation, the original sample is randomly partitioned into *K* subsamples. A single subsample is retained as the validation data for testing the model, and the remaining *K*-1 subsamples are used as training data. The cross-validation process is then repeated *K* times (folds). In our experiment, we take 5-fold cross-validation.

- **Mean Absolute Error (MAE)**

Since the output of prediction models estimate the probability the user will click ads in a specific ad session, if the result we predict is consist with true ad click probability, then we can say this prediction is perfect. To estimate the performance of each method, we calculate the deviation between the predicted probability and the true ad click probability which is either one or zero. The equation is defined as,

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|C_i - y'_i| \tag{14}$$

where n is the size of the testing data set, $C_i = 1$if it is a click and 0 otherwise. y' is the predicted probabilities in the range [0, 1].

➢ **Correlation Coefficients**

Pearson's correlation $\rho_{y',C}$ between the predicted probability and true probability(0 or 1) across all ad sessions in the test data is,

$$\rho_{y',C} = \frac{\sum_{i=1}^{n} C_i * y'_i - n * u_C * u_{y'}}{(n-1)\sigma_C * \sigma_{y'}} \quad (15)$$

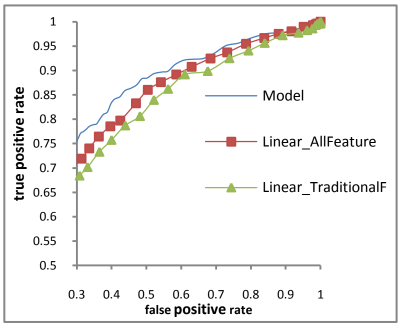
where σ is standard deviation and μ is the observed sample mean. This correlation coefficient is useful for detecting the presence of informative predictions, even in the presence of shifting and scaling. Thus we also consider it as one of our evaluation metrics.

4.2.3 Results

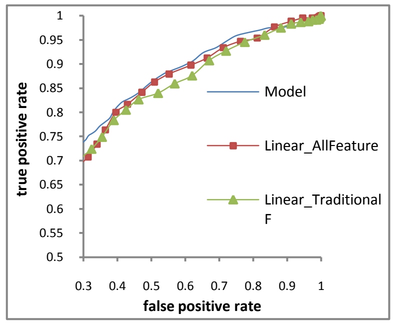
4.2.3.1 ROC Curve

We obtain 3 ROC curves in each round during 5-fold cross-validation, which are shown in Figure 4. In the figures, "Model" stands for the combined model described in Section 3. "Linear_AllFeature" is the linear regression using all feaures introduced in this paper. "Linear_TraditionalF" is the linear regression using traditional features. We treat the attitude features and commercial features as traditional ones because they have been mentionded in previous studies [30, 31].

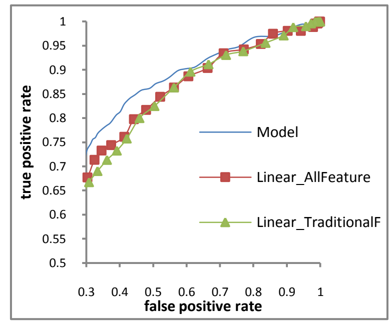
From the figure, we can see that the result of linear regression using all features is a little better than that using only tradictional features, which means the new features proposed in our study are helpful for ad click prediciton. The best ROC curve is the combined model, which means building the combined model is helpful for predicting ad click. Since we got the same consistent observation from the 5 runs, we say that the conclusion is stable with the changing of testing data.



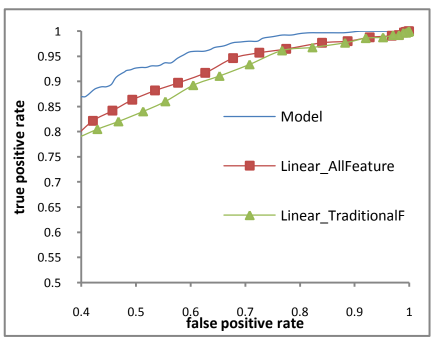
(a)Round I



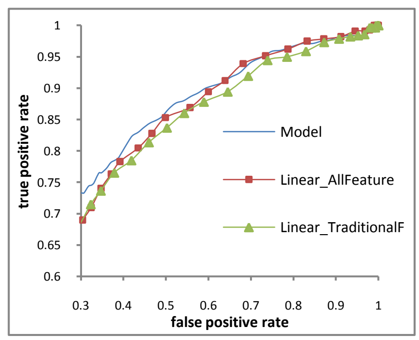
(b) Round II



(c) Round III



(d) Round IV



(e) Round V

Figure 4. ROC curve of the five runs

4.2.3.2 Mean Absolute Error

For the 5-fold cross validation, we calculate the Mean Absolute Error (MAE) for each round. The combined model used in experiments is introduced in Section 4. The average and standard deviation of MAE for each method is shown in Table 9.

Table 9. Comparison of MAE

Method	MAE	
	Average	Deviation
Model	0.120228	0.022456
Linear_AllFeature	0.263524	0.007934
Linear_TraditionalF	0.265664	0.010424

From Table 9 we can see that the average MAE of the combined model is much smaller than other methods, which means the combined model is a better method to approximate the probability of whether the user will click ads in a specific ad session.

4.2.3.3 Correlation Coefficients

For the 5-cross validation, we calculate the average and standard deviation of Pearson's correlation for each method, which is shown in Table 10.

Table 10. Comparison of Correlation

Method	Correlation	
	Average	Deviation
Model	0.297312	0.043769
Linear_AllFeature	0.264614	0.026988
Linear_TraditionalF	0.228738	0.046053

From Table 10, we can see that the average correlation of the combined model is the largest, which means the prediction power of the combined model is best compatible with the true probability. The linear regression using all features is a little worse, but better than the linear regression using only traditional features. The difference between the methods is not so salient compared with that of MAE, and this is because correlation calculation alleviates the effect of shifting and scaling.

4.2.4 *Sensitivity Analysis*

There are several parameters used in the experiments such as the window sizes *S*, enhance-factor *EN* and forget-factor *FOR* for habit learning, that have not been analyzed. In this subsection, we aim to analyze the sensitivity of results to the changing of these parameters.

To estimate the sensitivity of the parameters, we calculate the Pearson's correlation between the corresponding feature and ad click (0 or 1) across all ad sessions in the test data. Through tuning the parameter, we can see the sensitivity of the feature.

The window size should be balanced from 1 to the total number of time stamps *T* in our dataset. The sensitivity estimate is shown in Figure 5. From the figure we can see that the result is not so sensitive to the window size. In our experiment, we set the window size of Entropy to 2 and Wratio to 3, since it is a little better than other window sizes, which we can observe in Figure 5.

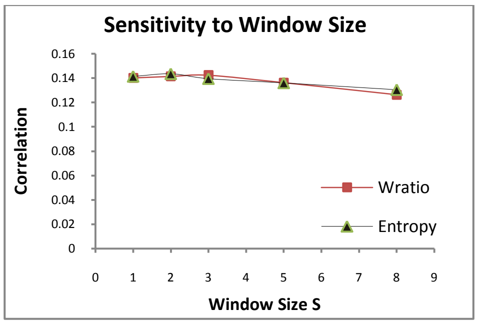


Figure 5. Sensitivity analysis to Window Size

To estimate the performance of enhance-factor *EN* and forget-factor *FOR*, we fix *FOR*=1.05 because the performance decrease a little when we decrease *FOR*. To see the effect when we increase *EN*, we calculate the correlation between the feature value and the ad click. The result is shown in Figure 6. From the figure we can see that it is generally not sensitive to the enhance-factor parameter. However *EN*=1.07 seems could be the best in Figure 6. In this case *EN* is a little larger than *FOR*. This is because for most of users, they do not click ads frequently and it's better to enhance the effect of ad click in this situation.

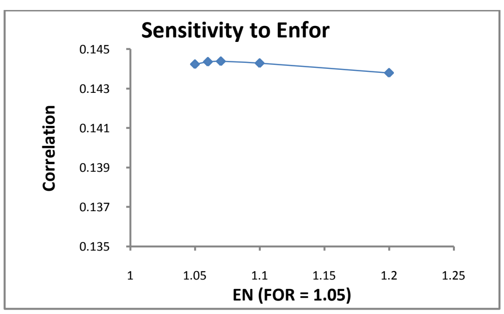


Figure 6. Sensitivity analysis of *EN* when *FOR*=1.05

5. RELATED WORK

Judith et al. [27] proposed to identify the mechanisms through which user's past behavior is linked to future behaviors. They suggested that in some domains, habit plays a dominant role in affecting user behaviors. There are a number of models proposed to predict human behaviors, such as Theory of Reasoned Action (TRA) and Theory of Planned Behavior (TPB). As an example application, Moez et al. [25] developed the Habit-Intention Model for users' IT usage behavior prediction. It is inspired by Triandis' suggestion [38] that habit can have both a direct and interactive effect on user behaviors. In their model, the factors affecting IT usage are attitude, habit etc. All these studies provide us good insights for understanding users' ad click behaviors. However, almost all of these studies are accomplished through subjective questionnaires.

In the field of Web data mining, the study taken by Jean-Louis et al. [7] estimated the different factors' effects on the Click-Through Rate (CTR) of banner ads. Adam et al. [35] examined the effect of banner style, including size and orientation, on interaction and CTRs. Patrali et al. [9] aimed to analyze the click proneness of repeated exposures to banner ads for different consumers in different situations. For sponsored search, Bernard et al. [22] examined the factors that influence users' bias against sponsored links. But their experiments are taken under controlled conditions through demographic questionnaires.

For the ad centered research, Matthew et al. [31] used features of ads such as terms in ad title and advertisers' bid keywords to learn a model that predicts the click-through rate for new ads. Moira et al. [30] proposed a method that used the historical CTR of related terms to estimate the rate for low-volume or novel terms. For the user centered research, Yan et al. [41] validated that Behavior Targeting (BT) can truly help online advertising in search engines. To our knowledge, the work in [26] is the first study that uses search toolbar data to estimate the effectiveness of sponsored search advertising. The work in [20] focused on the evidences that affect Google's paid search result rankings. The work in [23] utilized click-through logs of a Meta search engine for analyzing user preference on sponsored search results. However, the reasons what are leading users to click ads was not discussed in these works, which is the primary motivation for our work.

6. CONCLUSION AND FUTURE WORK

We solve the problem by considering user habit, attitude and commercial intents. The empirical study results state that features extracted for these three factors have strong correlations with users' ad click behaviors. By the experiments, these proposed features together with the prediction model both assist the classical machine learning algorithms to perform better. This work focuses on improving the effect of these features, instead of extracting as many features as possible. Thus, in future research, we plan to mine more features to promote the prediction model.

7. REFERENCES

[1] Ajzen, I. Attitudes, personality, and behavior. In Chicago: Dorsey Press, 1988.

[2] Ajzen, I. The Theory of Planned Behavior. Organizational Behavior and Human Decision Processes, Vol. 50, pp. 179-211, 1991.

[3] Allport, G. W. Attitudes. In C. Murchison (Ed.), A handbook of social psychology (pp. 798-844). Worcester, MA: Clark University Press, 1935.

[4] Andrews, D. A robust method for multiple linear regression. Technometrics1,6 , 523--531, 1974.

[5] Breiman, L., Friedman, J., Olshen, R. A. and Stone, C. J. Classification and regression trees. In Monterey, CA: Wadsworth, 1984.

[6] Campbell, D. T. Social attitudes and other acquired behavioral dispositions. In S. Koch (Ed.), Psychology: A study of a science. (Vol. 6, pp. 94-172). New York: McGraw-Hill, 1963.

[7] Chandon, J. and Chtourou, M. Factors affecting click through rates. In Society of Consumer Psychology Conference. Seattle, 2001.
[8] Chapelle, O., Zhang, Y. A dynamic bayesian network click model for web search ranking. In Proceedings of WWW'2009. pp.1~10, 2009.
[9] Chatterjee, P., Hoffman. D. L. and Novak. T. P. Modeling the Clickstream: Implications for Web-Based Advertising Efforts. In Marketing Science22 (4), (Autumn, 2003), 520-541, 2003.
[10] Chin, W. The Partial Least Square Approach for Structural Equation Modeling. In Modern Methods for Business Research, pp. 295-336, 1998.
[11] Chin, W. and Gopal, A. Adoption intention in GSS:Relative Importance of Beliefs. In DATABASE for Advances in Information System(26:2&3), pp.42-64, 1995.
[12] Compeau, D. and Higgins, C. Application of social cognitive theory to training for computer skills. Information Systems Research, 6(2), 118-143, 1995.
[13] Craswell, N., Zoeter, O., Taylor, M., and Ramsey, B. An experimental comparison of click position-bias models. In WSDM '08: Proceedings of the international conference on Web search and web data mining, pages 87–94. ACM, 2008.
[14] Dai, H. K., Zhao, L., Nie, Z., Wen, J.-R., Wang, L. and Li, Y. Detecting online commercial intention (oci). In WWW 2006, 2006.
[15] Dupret, G. and Piwowarski, B. User browsing model to predict search engine click data from past observations. In SIGIR 08: Proceedings of the 31st Annual International Conference on Research and Development in Information Retrieval, 2008.
[16] Eagly, A. and Chaiken, S. The psychology of attitudes. In Fort Worth,TX:Harcourt Brace Jovanovich, 1993.
[17] Feng, J., Bhargava, H. K. and Pennock, D. M. Implementing sponsored search in web search engines: Computational evaluation of alternative mechanisms. In INFORMS Journal on Computing, 2005.
[18] Fishbein, M., and Ajzen, I. Belief, attitude, intention, and behavior: An introduction to theory and research. In Reading, MA: Addision-Wesley, 1975.
[19] Fornell, C. and Bookstein, F. Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. In Journal of Marketing Research, 19, 440–452, 1982.
[20] Ghose, A. and Yang S. An empirical analysis of sponsored search performance in search engine advertising. In Proceedings of the international Conference on Web Search and Web Data Mining (Palo Alto, California, USA, February 11 - 12, 2008). WSDM '08. ACM, New York, NY, 241-250, 2008.
[21] Hotchkiss, G. Inside the mind of the searcher. In Retrieved March 15, 2005.
[22] Jansen, B. and Resnick, M. An examination of searchers' perceptions of non-sponsored and sponsored links during ecommerce Web searching. In J. Amer. Soc. Inform. Sci. Technol. 57, 1949–1961, 2006.
[23] Jansen, B. and Spink, A. The effect on click-through of combining sponsored and non-sponsored search engine results in a single list. In Sponsored Links Workshop. WWW 2007: 16th International World Wide Web Conference, May 8-12, Banff, Canada, 2007.
[24] Kumar, S., Dawande, M., and Mookerjee, V. S. Revenue maximization in web advertisement. In Proceedings of the Thirteenth Annual Workshop on Information Technologies And Systems (WITS'03). Seattle, WA, 121–126, 2003.
[25] Limayem, M., Hirt, S., Chin, W. Intention does not always matter: the contingent role of habit on IT usage behavior. In European Conference on Information Systems, Bled, Slovenia, pp.27-9, 2001.
[26] Liu, Y., Zhou, B., Ma, S. Effectiveness of Online Sponsored Search Based on User Log Analysis. In Journal of Computational Information Systems. Vol.5(3): pp. 1175-1184, 2009.
[27] Ouelette, J. and Wood W. Habit and intention in everyday life: the multiple processes by which past behavior predicts future behavior. In Psychological Bulletin. v124 i1. 54-74, 1998.
[28] Quinlan, J. C4.5: Programs for machine learning. In San Francisco, CA: Morgan Kaufmann, 1993.
[29] Radlinski, F., Broder, A., Ciccolo, P., Gabrilovich, E., Josifovski, V., and Riedel, L. Optimizing relevance and revenue in ad search: a query substitution approach. In SIGIR '08, 2008.
[30] Regelson, M. and Fain, D. Predicting click-through rate using keyword clusters, In Proceedings of the Second Workshop on Sponsored Search Auctions, 2006.
[31] Richardson, M., Dominowska, E., Ragno, R. Predicting clicks: estimating the click-through rate for new ads. In Proceedings of the 16th International Conference on World Wide Web (WWW 2007), pp. 521–530. ACM, New York , 2007.
[32] Ronis, D. L., Yates J. F., and Kirscht J. P. Attitudes, decisions, and habits as determinants of repeated behavior. In A. R. Pratkanis, S. J.Breckler, & A. G. Greenwald (Eds.), Attitude structure and function (pp.213-239). Hillsdale, NJ: Erlbaum, 1989.
[33] Sherman, S. J., & Fazio, R. H. Parallels between attitudes and traits as predictors of behavior. In Journal of Personality, 51, 308—345, 1983.
[34] Shi, H. Best-first Decision Tree Learning. In The University of Waikato. 2007.
[35] Sigel,A., Braun, Sena, G. M. The Impact Of Banner Ad Styles On Interaction And Click-Through Rates. In VOL IX,No. 2, 2008.
[36] Tang, Y. The Effects of Commercial Intention, Humor, and Media Richness on Receivers' Forward Intention in Internet Viral Marketing Campaigns. In Information Management, 2005.
[37] Tenenhaus, M., Vinzia, V. PLS path modeling. In Computational Statistics & Data Analysis, Vol. 48, S. 159-205, 2004.
[38] Triandis, H. Values, attitudes, and interpersonal behavior. In H.E. Howe (Ed.), Nebraska Symposium on Motivation, 1979: Beliefs, Attitudes and Values. Lincoln: University of Nebraska Press, pp. 195-259, 1980.
[39] Weber, T.A., Zheng, Z. E. A model of search intermediaries and paid referrals. In Tech. rep.,The Wharton School, University of Pennsylvania, 2003.
[40] Wold, H. Introduction to the Second Generation of Multivariate Analysis. In Theoretical Empiricism, H. Wold(ed.), Paragon House, New York, pp.vii-xi, 1989.
[41] Yan, J., Liu N. et al. How much can Behavioral Targeting Help Online Advertising? In Proceedings of the 18th International World Wide Web Conference (WWW), 2009.
[42] Zweig, M., Campbell, G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. In Clin Chem. 1993;39:1589, 1993.