Probabilistic Latent Semantic User Segmentation for Behavioral Targeted Advertising*

Xiaohui Wu[1,2], Jun Yan[2], Ning Liu[2], Shuicheng Yan[3], Ying Chen[1], Zheng Chen[2]

[1]Department of Computer Science
Beijing Institute of Technology
Beijing, China, 100081
xiaohuiwu85@gmail.com, chenying1@bit.edu.cn

[2]Microsoft Research Asia
Sigma Center, 49 Zhichun Road
Beijing, China, 100080
{v-xwu, junyan, ningl, zhengc}@microsoft.com

[3]National University of Singapore
Office E4-05-11, 4 Engineering Drive 3
Singapore, 117576
eleyans@nus.edu.sg

ABSTRACT

Behavioral Targeting (BT), which aims to deliver the most appropriate advertisements to the most appropriate users, is attracting much attention in online advertising market. A key challenge of BT is how to automatically segment users for ads delivery, and good user segmentation may significantly improve the ad click-through rate (CTR). Different from classical user segmentation strategies, which rarely take the semantics of user behaviors into consideration, we propose in this paper a novel user segmentation algorithm named Probabilistic Latent Semantic User Segmentation (PLSUS). PLSUS adopts the probabilistic latent semantic analysis to mine the relationship between users and their behaviors so as to segment users in a semantic manner. We perform experiments on the real world ad click through log of a commercial search engine. Comparing with the other two classical clustering algorithms, K-Means and CLUTO, PLSUS can further improve the ads CTR up to 100%. To our best knowledge, this work is an early semantic user segmentation study for BT in academia.

Categories and Subject Descriptors

H.3.5 **[Information Storage and Retrieval]**: Online Information Service – *Commercial Service*; I.5.1 **[Pattern Recognition]**: Models – *Statistical*

General Terms

Algorithms, Performance, Experimentation

Keywords

Behavioral Targeting (BT), User segmentation, probabilistic latent semantic analysis

1. INTRODUCTION

Nowadays, a large number of advertisers would like to publish their advertisements through Internet, which brought a new developing field known as online advertising science. Sponsored search [9] and contextual ads [4] are two of the most widely studied online advertising business models. Besides, Behavioral Targeting, which aims to analyze users' behaviors to deliver appropriate ads to potential consumers, has been validated to make online advertising more effective [19]. A crucial portion in BT is the problem of user segmentation, which aims at grouping users into user segments with similar behaviors. Since advertisers generally select user segments most relevant to their ads, if users with similar purchase intentions are successfully gathered into the same segment, advertiser may gain more profit from the ads delivery. Thus, the quality of user segmentation has dominant impact on the performance of behavioral targeted advertising.

In this paper, we focus on the problem of user segmentation for BT in search engine advertising. User segmentation is a process arranging each user into one or more segments to guarantee that users with similar interests or purchase intentions are within the same segment. We formulate the problem of user segmentation as follows. Suppose a set of online users is given. For each user, we adopt his/her historical online behaviors such as queries to depict his/her interests. Some ads have been displayed to these users, and these ads are recorded with the status whether they are clicked in the impression. Our objective is to group all users into appropriate segments by the analysis of user behaviors in order to improve the ad click probability within the user segments in contrast to the massive market ads .

Conventional user segmentation approaches such as classification and clustering stand two limitations. (1) Many traditional strategies utilize keywords as features, and then implement clustering or classification on these features. In this way, two users, who have the similar buying intentions but have no common words between each other, shall not be put into the same segment. (2) Many classical clustering methods do not allow an object to belong to multiple clusters, which means one user can only stay in a unique segment. We notice that semantic approaches such as Latent Semantic Analysis (LSA) [8], Probabilistic Latent Semantic Analysis (PLSA) [13] and Latent Dirichlet Allocation (LDA) [1] are widely studied and adopted in field of document classification. Among those approaches, PLSA effectively mines the relationship between document and word with a hidden variable called topic. Besides, PLSA has the ability

ADKDD'09, June 28, 2009, Paris, France.

*The work was done when the first author was intern at Microsoft Research Asia.

to take one document into multiple topics. Motivated by PLSA in the field of text mining, we propose to analyze the similarity between user-query and document-word and present a semantic approach called Probabilistic Latent Semantic User Segmentation (PLSUS) for handling the limitations of those traditional user segmentation strategies.

In order to execute semantic user segmentation for BT, we firstly split users' search queries into terms and thus each user can be represented as a collection of terms [19]. Using this Bag of Words (BOW) representation [18], the latent variable, which presents the topics, i.e. semantic user interests, is involved to represent users. This topical variable is able to bridge users and their observed behaviors. We will show that the latent semantic topics can present users' interests implying potential purchase intention in our experiments. For this reason, we directly utilize the latent semantic topics to segment users. The Expectation Maximization (EM) approach is applied to mine the latent semantic topics. Since a user may have multiple interests, to better image the user's interests, we set a threshold and push the user into those segments with the probabilities larger than the predefined threshold.

In the experiments, we compare our proposed PLSUS and a modified version, which is known as Single-PLSUS with two commonly used clustering algorithms, CLUTO and k-Means. Single-PLSUS only allows a user in a unique segment as many traditional clustering algorithms do. The results show that PLSUS can improve the ads CTR up to 100%. In addition, PLSUS has good performance on classical F-measure.

The rest of this paper is organized as follows. In Section 2, we introduce the background knowledge about BT and semantic graphical models in text mining. In Section 3, we describe our solution to semantic user segmentation, namely PLSUS. In Section 4, we introduce the experimental configuration and results with analysis. Finally in Section 5, we conclude this paper along with future work discussion.

2. BACKGROUND

In this section, we introduce the background knowledge for better understanding this work. We demonstrate the basic knowledge on BT including the definition and related commercial systems in Section 2.1. In addition, we review the semantic approaches such as LSA, PLSA, and LDA in Section 2.2.

2.1 Behavior Targeting

Behavioral Targeting is an advertising methodology, which is burgeoning in online advertising. With this technique, ads can be effectively delivered to the most relevant users. Behavior Targeting may improve the performance of online advertisement delivery by two major steps, namely, user segmentation and user segments ranking. In the user segmentation step, based on online behavior such as visited websites, clicked pages and input queries, users are located into some user segments created in system. In the user segments ranking step, given an ad, user segments are ranked by relevance and the top segments are chosen for ads delivery. Thus, BT successfully displays the ads to those most appropriate users.

At present, BT is attracting more and more attention in both industry and academia. In industry, a large amount of commercial systems involving Behavioral Targeting were proposed: Adlink [20], which takes the short user session into consideration for BT, DoubleClick [24], which adopts special features such as browser types and operation systems of users to improve the user segmentation step, Specificmedia [28], which predicts each user's interest and purchase intention as a score, and the Yahoo! Smart ads [30], which integrates the demographic and geographic targeting. Additionally, Almond Net [21], Blue Lithium [27], Burst [23], NebuAd [25], Phorm [26], Revenue Science [26], and TACODA [29] are the commercial systems with BT. In academia, Yan et al. [19] first studied the improvement of BT in commercial search engines from three aspects including effectiveness, improvement, and the best strategy for BT.

User segmentation is a process arranging each user into one or more segments by a specific criterion. In BT, this criterion is to endeavor to guarantee that users with similar interests and purchase intentions are in the same segment. However, we cannot derive that information directly. The most widespread way is mining the user behaviors to represent user interests and purchase intentions. That means users with the similar behaviors imply that they have the similar favors. Thus, user segmentation for BT can be described as attempting to place each user in one or more segments for guaranteeing that the users with similar behaviors are in the same segment. Since advertisers tend to choose most relevant segments to pay, the quality of user segmentation is extraordinarily crucial. On one hand, if system can gather more users with similar interests into one segment, advertisers will buy fewer segments to deliver their ads. On the other hand, apparently, CTR is to improve if the similarity between each pair of users within the same segment is large. Thus, user segmentation is a key problem in BT application.

Traditional user segmentation approaches for BT can be classified into three categories, namely manual user segmentation, user classification, and user clustering. Manual rule based user segmentation, which classifies users into segments manually, suffers from a significant deficiency in time cost. As a result of that large scale data is used for BT, this method was hardly adopted by the commercial systems. User classification and user segmentation respectively implement classification and clustering for users. The traditional clustering or classification approaches have two limitations in this application scenario. (1) Users are segmented only based on 'contents' of their behaviors, not their semantic interests. With the Bag of Words model, traditional strategies utilize terms as features in order to implement clustering. That means two users with the similar purchase intentions but without same terms between each other have little chance to be grouped into one segment. (2) Many clustering methods which are widely used for BT concentrate on settling one object in one cluster. On account of this limitation, if a user has two completely different interests, only one interest can be presented and the other one has to be discarded. Thus, it is desired to propose new semantic segmentation approaches for BT.

2.2 Semantic Analysis

Semantic analysis, which is a well established technique in industry, mines hidden semantic relationships among objects. Latent Semantic Analysis (LSA) [8] is the well-known approach for deriving the latent semantic relationship and widely used in automatic indexing and information retrieval. The main idea is mapping high-dimensional vectors to low-dimensional ones in the latent semantic space. Probabilistic Latent Semantic Analysis (PLSA) model [13, 15], which is derived from LSA, is able to capture hidden variables with solid statistical foundation. Each object is represented by the convex combination of 'topic', which is a latent variable in PLSA. Latent Dirichlet Allocation (LDA) [1] is similar to PLSA. The difference between these two models is that the topic distribution is assumed to have a Dirichlet prior in LDA. In document classification, LDA derives more reasonable mixtures of topics. However, the work in [11] has proved that the PLSA model is equivalent to the LDA model under a uniform Dirichlet prior distribution.

In this work, we focus on PLSA to derive our PLSUS model. PLSA is a significant breakthrough, since it can discover latent variables with more flexibility. Besides, using the EM algorithm, we can easily estimate the value in PLSA. In practice, PLSA is widely used in many fields such as document classification [2, 3, 10, 17], information retrieval [14], web usage mining [16], co-citation analysis [5, 6] and collaborative filtering [7, 12]. However, there are rare works which apply PLSA to user segmentation for BT. In our study, following the Bag of Words model, we describe each user as a collection of terms, which are extracted from their behaviors, such that we can represent users in the Bag of Words model, which is similar to the commonly used document representation strategy.

3. PROBABILISTIC LATENT SEMANTIC USER SEGMENTATION (PLSUS)

In this section, we introduce our semantic user segmentation algorithm. PLSA, which can discover the latent relationship between two objects, is widely studied in document classification and clustering problems. In text mining, we generally use the Bag of Words model [18] to represent documents. According to the work of Yan et al. [19], users' behaviors can be represented by their historical queries. Notice the fact that query consists of terms, thus we can treat each query as one set of terms. Through this way, each user can be represented by a bag of words, which is the same as the representation of text document. Let $u_i \in U = \{u_1, u_2, ..., u_n\}$ stand for a user, where U presents the set of all users for BT, suppose $t_j \in T = \{t_1, t_2, ..., t_m\}$ is a term, where T represents the vocabulary of all terms used by all users. We define T_{u_i} as the set of all terms used by u_i, thus,

$$T = \bigcup_{u_i \in U} T_{u_i}$$

Then, we define the co-occurrence matrix $N = \{n(u_i, t_j)\}$, where $n(u_i, t_j)$ describes the number of time t_j used by u_i.

To semantically segment users, we introduce the latent variable $z_k \in Z = \{z_1, z_2, ..., z_l\}$ which represents the topics, i.e. semantic intentions of users. This latent variable has the close relationship with both user and query, which has been transformed into terms. From the user's perspective, topic implies the hidden interest of user. On the other hand, from the term's perspective, terms in one topic may be gathered with some specified field. Here, we assume that for a given topic variable z_k, users and terms are independent to each other. We adopt the classical aspect model [13] in PLSUS. The graphic model of aspect model is given by Figure 1.

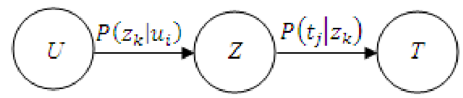

Figure 1. Graph of the aspect model

In the BT scenario, each user has the probability $P(z_k \mid u_i)$ to generate a topic z_k, and then z_k has the probability $P(t_j \mid z_k)$ to generate term t_j. Given the basic model,

$$P(u_i, t_j) = P(u_i)P(t_j \mid u_i)$$

$$P(t_j \mid u_i) = \sum_{z_k \in Z} P(t_j \mid z_k)P(z_k \mid u_i)$$

Notice that, this model contains the probability $P(z_k \mid u_i)$ and $P(u_i)$ which are not convenient to compute. Thus, we transform this model into another equaling form,

$$P(u_i, t_j) = \sum_{z_k \in Z} P(z_k)P(u_i \mid z_k)P(t_j \mid z_k),$$

where $P(z_k)$ presents the probability that z_k is observed in Z, $P(u_i \mid z_k)$ is the probability that u_i is relevant to the given topic z_k and $P(t_j \mid z_k)$ is the probability that t_j is related to the given topic z_k. The Graphical model representation is shown in Figure 2.

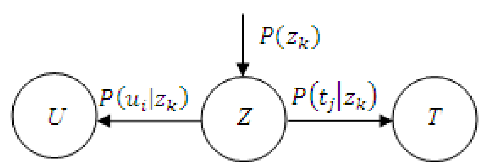

Figure 2. Graph of the PLSUS.

The same as PLSA in the field of text mining, we aim to maximize the likelihood defined as,

$$L = \sum_{i=1}^{n}\sum_{j=1}^{m} n(u_i, t_j)\log P(u_i, t_j)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} n(u_i, t_j)\log \sum_{k=1}^{l} P(z_k)P(u_i \mid z_k)P(t_j \mid z_k)$$

In order to maximize L, we adapt the classical Expectation Maximization (EM) approach. EM approach is widely used in computing maximum likelihood in latent variable model. EM is an iterative method which alternates between performing two steps. (1) Expectation step (E step). Using the current estimates of parameters, we compute the posterior probabilities $P(z_k \mid u_i, t_j)$ for the latent variable. (2) Maximization step (M step). Aiming to maximize complete maximize likelihood $E[L^c]$, we update $P(z_k)$, $P(u_i \mid z_k)$ and $P(t_j \mid z_k)$.

After finishing EM computation, PLSUS aims to segment users with the model obtained. Since the topic has the close relationship with user and term, apparently, topic can be used as user segment. In this way, the semantic attributes become the dominant factors in user segmentation. Thus, we aim to solve the question of how to segment users into different topics. To solve this question, we focus on an important probability $P(z_k \mid u_i)$ which presents the topic (user segment) z_k is observed with a given user u_i. It can describe how close the relationship between z_k and u_i is. $P(z_k \mid u_i)$ is able to be computed by,

$$P(z_k \mid u_i) = \frac{\sum_{j=1}^{m} n(u_i, t_j) P(z_k \mid u_i, t_j)}{\sum_{j=1}^{m} \sum_{k'=1}^{l} n(u_i, t_j) P(z_{k'} \mid u_i, t_j)}.$$

Intuitively, the easiest way to segment users into topics is that, computing all $P(z_k \mid u_i)$, $z_k \in Z$ for each u_i, and then putting u_i into the topic with the highest $P(z_k \mid u_i)$. However, this approach of user segmentation cannot handle the following circumstance: If a user is interested in sports and cooking while there are two topics which exactly imply sports and cooking, this segmentation method will choose only one topic for a user at most. In this way, we may lose a user's interest. In order to get over this deficiency, we present a novel approach for segmenting the users based on the probability $P(z_k \mid u_i)$. Here, we apply a *threshold* for user segmentation. Let S be the set of user segments and $s_k \in S$ as the segment with topic z_k, thus the user segmentation approach is,

$$\begin{array}{ll} u_i \in s_k & if\ P(z_k \mid u_i) > threshold \\ u_i \notin s_k & \text{otherwise} \end{array}$$

Comparing with those traditional clustering methods, this simple method allows one user belong to multiple segments.

4. EXPERIMENTS

In this section, we systematically evaluate the proposed PLSUS algorithm. Two normal clustering methods are used as baselines in experiments. Also, to better compare with normal clustering approaches, a modified PLSUS which we called Single-PLSUS is inducted. Some evaluations are used in our experiments to measure the performance of each approach.

4.1 Data Sets

In this part, we use a one day's ads click-through log record collecting from a commercial search engine. This data can effectively present users' click-through behaviors. Table 1 shows the format of this data used in our experiments. From this table, we can see that there are four properties for the data we focused on. UserId presents a specified user, different user has different UserId. Similar to UserId, AdId is used as the unique identification for each advertisement. Query shows the content of a query used by user, and we can divide it into terms to adapt to PLSUS. ClickCnt is an important property which is used in our evaluation metrics such as CTR. From the example in Table 1, we know a specified user with UserId EEEC97C25FD50C1AB282 D39FB13976D9 used a query whose content is "books", and then the system displays an advertisement with AdId 3238034 to this user. However, this user did not click this ad.

Table 1. Format of log record used in our experiments.

UserId	EEEC97C25FD50C1AB282D39FB13976D9
Query	Books
AdId	3238034
ClickCnt	0

We use two datasets including 120,000 and 150,000 log records respectively to verify the performance of PLSUS. Both of them contain thousands of users. In our experiments, we take all users in 120,000 log dataset into 5 and 10 segments, while all users in 150,000 log dataset are pushed in 10 and 20 segments respectively using different approaches.

4.2 Experiment Setup

In this part, we introduce the key steps of our experiments. In user segmentation, let $A = \{a_1, a_2, ..., a_n\}$ be the set of ads in our dataset, $U_i = \{u_{i1}, u_{i2}, ..., u_{im}\}$ be the group of users who have displayed a_i. Furthermore, after we segment users with different approaches, we acquire the user segments. Thus, we define user $D(U_i) = \{d_1(U_i), d_2(U_i), ..., d_k(U_i)\}$, $i = 1, 2, ..., n$ be the distribution of U_i with our obtained user segments and $d_k(U_i)$ the set of users who belong to the *kth* segment. Apparently, the *kth* segment can be describe as,

$$d_k \bigcup_{i=1,2,\ldots,n} d_k(U_i)$$

The key steps in our experiments are,

(1) We compare PLSUS, Single-PLSUS, k-Means and CLUTO in our dataset, where Single-PLSUS is a modified PLSUS which we will introduce in latter section of this paper.

(2) We utilize the different *threshold* which is adopted in segmenting users after coming out the final model by EM algorithm to test the sensitivity of PLSUS.

4.3 Evaluation Metrics

In [19], Yan introduced some evaluations which can measure the BT's performance effectively. Consulting these good evaluations, we perform four evaluations to measure the performance of each approach and to compare our solution with the baselines. They are, ads Click-Through Rate (CTR), ads Click-Through Rate Improvement, ads click Entropy and F-measure.

With the symbols we defined, CTR can be represented by,

$$CTR(a_i) = \frac{1}{m} \sum_{j=1}^{m_i} \delta(u_{ij}),$$

where $\delta(u_{ij})$ is defined as,

$$\delta(u_{ij}) = \begin{cases} 1 & \text{if } u_{ij} \text{clicked} a_i \\ 0 & \text{otherwise} \end{cases}$$

CTR of a_i over user segment d_k is,

$$CTR(a_i \mid d_k) = \frac{1}{|d_k(U_i)|} \sum_{u_{ij} \in d_k(U_i)} \delta(u_{ij}),$$

where $|d_k(U_i)|$ is the number of users in $d_k(U_i)$.

Note that $CTR(a_i)$ is the raw CTR. in other words, $CTR(a_i)$ is the CTR over all users displayed a_i. $CTR(a_i \mid d_k)$ presents the CTR of each user segment after segmentation.

In order to measure the improvement of CTR by user segmentation, we define a new evaluation metric for PLSUS. This new evaluation should satisfy two conditions,

(1) Maximum: choosing the segment which has maximum CTR. This is reasonable because ad publisher would like to recommend the user segment with highest ad click probability to advertiser for ads delivery.

(2) Majority: the number of users in this segment cannot be less than average. This condition can reduce some special situation. For example, the *kth* user segment only has 1 user and he/she clicked a_i. Then, $CTR(a_i \mid d_k) = 1$. Apparently, this segment is not appropriate to be recommended to advertiser.

Integrating these two conditions, we define the CTR improvement for a_i as,

$$\Delta(a_i) = \frac{CTR(a_i \mid d^*(a_i)) - CTR(a_i)}{CTR(a_i)},$$

where

$$d^*(a_i) = \arg\max\{CTR(a_i \mid d_k), d_k \in \tilde{d}\}$$

$$\tilde{d} = \{d_k \mid k = 1,2,..., K \text{ and } \frac{|d_k(U_i)|}{m_i} \geq \frac{1}{K}\}$$

Thus, CTR improvement $\Delta = \sum_i \Delta(a_i) / n$.

Entropy is defined as,

$$Enp(a_i) = -\sum_{k=1}^{K} P(d_k \mid a_i) \log P(d_k \mid a_i),$$

where,

$$P(d_k \mid a_i) = \frac{1}{m_i} \sum_{u_{ij} \in d_k(U_i)} \delta(u_{ij})$$

Note that the smaller the Entropy is, the better results we will obtain [19].

The classical F-measure including Precision, Recall and F measure, are defined as,

$$Pre(a_i \mid d_k) = CTR(a_i \mid d_k)$$

$$Rec(a_i \mid d_k) = \frac{\sum_{u_{ij} \in d_k(U_i)} \delta(u_{ij})}{\sum_{j=1}^{m_i} \delta(u_{ij})}$$

$$F(a_i \mid d_k) = \frac{2Pre(a_i \mid d_k) Rec(a_i \mid d_k)}{Pre(a_i \mid d_k) + Rec(a_i \mid d_k)}$$

where the larger F-measure is, the better performance we have.

4.4 Results

In this part, we introduce the details in our experiments and show the results. To show the performance of PLSUS, we aim to compare PLSUS with traditional clustering methods. CLUTO and k-Means are selected as the baselines. However, it is unfair to compare CLUTO and k-Means with PLSUS since PLSUS allows one user belong to multiple segments, while both CLUTO and k-Means permit one user to belong to only one user segment. In order to solve this problem, a Single-PLSUS is implemented to bridge the gap between PLSUS and traditional clustering approaches. By Single-PLSUS, a given user u_i is settled in a unique segment z_k which has the max $P(z_k \mid u_i)$. On one hand, comparing Single-PLSUS with CLUTO and k-Means can show whether the semantic approach improves BT's performance. On the other hand, it can clearly show the impact on allowing one user to belong to multiple segments by comparison between PLSUS and Single-PLSUS. The results are shown in Table 2-4. Note that the best results are in bold face. Note that we set $threshold = 0.2$, the further explanation is shown in the latter sections.

CTR is one of the most basic and critical evaluation metric for online advertising problems. From the Table 2, we can generally observe two phenomena. First, by increasing the number of segments, the improvement of CTR is increasing simultaneously. In the 150,000 log dataset, as the segments doubled, the improvement of CTR increases two fold. In the same dataset, with the 20 segments, the PLSUS improves CTR up to 100% against traditional CLUTO. Second, all semantic approaches have the good performances on CTR improvement. By further analysis, Single-PLSUS totally exceeds CLUTO and k-Means. This fact proves that the semantic approach is appropriate to be adopted in BT. Since we gathered all queries used for each user and divide these queries into terms, we discover the correspondence between user-query and document-words. The results verify the correctness of our idea. The observation of comparison between PLSUS and Single-PLSUS shows the advantages from allowing user to be pushed into multiple segments. Besides, in Yan's work [19], CTR improvement with CLUTO and k-Means are around 100% by group users into 20 segments, which has been proved by our experimental results. Since Yan's experiments shown that CTR improvement can reach to 670% by 160 user segments in the large scale dataset, we are confident to expect that we can improve CTR more than that if we group users into more segments. In our future work, we will increase the scalability to verify this conclusion.

We compute the average ads click Entropy over all ads in the dataset we used. The result is shown in Table 3. Generally, all user segmentation approaches' entropies are almost the same. In this case, entropy has less effect on distinction among those methods than CTR. From the detailed observation, we discover that the entropy of PLSUS is larger than others. Considering their attributes, the reason is easy to get. The same criterion of user segmentation, which allows single user belong to multiple segments, is used in PLSUS. That means there is more than one segment which may have been delivered an ad many times. In this way, the entropy is naturally larger than those user segmentation approaches which only associate one user with one segment.

Table 2. CTR improvement of different user segmentation strategies.

	5 segments in 120,000 data	10 segments in 120,000 data	10 segments in 150,000 data	20 segments in 150,000 data
PLSUS	**0.7876**	**1.3583**	**1.3036**	**2.6549**
Single-PLSUS	0.6753	1.2985	1.2353	2.5287
CLUTO	0.6444	0.7399	0.7447	1.2076
k-Means	0.5440	0.7761	0.8616	1.0324

Table 3. Ads click Entropy of different user segmentation strategies.

	5 segments in 120,000 data	10 segments in 120,000 data	10 segments in 150,000 data	20 segments in 150,000 data
PLSUS	0.1636	0.1780	0.1824	0.1735
Single-PLSUS	0.1527	0.1542	0.1590	0.1611
CLUTO	**0.1506**	0.1586	**0.1531**	**0.1540**
k-Means	0.1532	**0.1515**	0.1575	0.1554

Table 4. F-measure of different user segmentation strategies.

		5 segments in 120,000 data	10 segments in 120,000 data	10 segments in 150,000 data	20 segments in 150,000 data
PLSUS	Precision	**0.9947%**	**1.0628%**	1.0954%	**1.2424%**
	Recall	1.3116%	1.3080%	1.3221%	1.3407%
	F	**1.0503%**	**1.1071%**	**1.1414%**	**1.2567%**
Single-PLSUS	Precision	0.9932%	1.0568%	**1.0972%**	1.2100%
	Recall	1.2672%	1.2927%	1.2896%	1.3410%
	F	1.0443%	1.1005%	1.1378%	1.2332%
CLUTO	Precision	0.9271%	0.9634%	0.9546%	1.0019%
	Recall	**1.3386%**	**1.3718%**	1.3824%	1.3979%
	F	0.9958%	1.0283%	1.0229%	1.0656%
k-Means	Precision	0.9196%	0.9197%	0.9663%	0.9833%
	Recall	1.3122%	1.3708%	**1.3930%**	**1.4083%**
	F	0.9870%	0.9945%	1.0346%	1.0520%

Precision, Recall and F-measure are shown in Table 4. Note that, the results reported in this table are the average over all ads. First of all, we discover the two facts that: (1) semantic approaches have better presentations in Precision. Since we choose the CTR as the Precision, this result can be predicted by CTR improvement. (2) Within three semantic approaches, PLSUS performs better than others. By these two facts, we can conclude that our proposed methods are helpful to improve the Precision (CTR). An interesting observation is the Recall of traditional clustering approaches is higher than others in our two small datasets. Considering the low precision, we can decide that the high-CTR segments clustered by CLUTO or k-Means should include many users. In other words, the way to improve CTR of a segment in traditional approaches is to add more users to this segment. On the contrary, semantic user segmentation can improve the CTR without building user segment with too large population. This characteristic is very useful for accurate ads delivery. Integrating Precision and Recall, the F-measure can evaluate the performance of user segmentation. From the results of high F-measures of PLSUS and Single-PLSUS, we can draw the conclusion that semantic user segmentation has better performance than classical clustering methods.

Finally, we discuss the influence of parameter *threshold* in PLSUS model. We set up a series of experiments which group users into 10 segments on the 120,000 log record data. Apparently, if $threshold \geq 0.5$, the output of PLSUS will be constant. Thus, we set *threshold* from 0.05 to 0.5 and the Figure 3-4 display the results. Since bigger *threshold* indicates that user have smaller chance to be collected into multiple segments, the CTR improvement lowers down when *threshold* becomes bigger in Figure 3. However, if we take a too small *threshold*, each user will have big opportunity to be settled in many segments. In this way, each segment will contain too much users and lead big entropy. The result in Figure 4 shows this fact. Analyzing Figure 3-4, we consider that *threshold* around 0.2 can perform good performance both on CTR improvement and entropy. Therefore, we set $threshold = 0.2$ for PLSUS in the experiment which compares four user segmentation approaches.

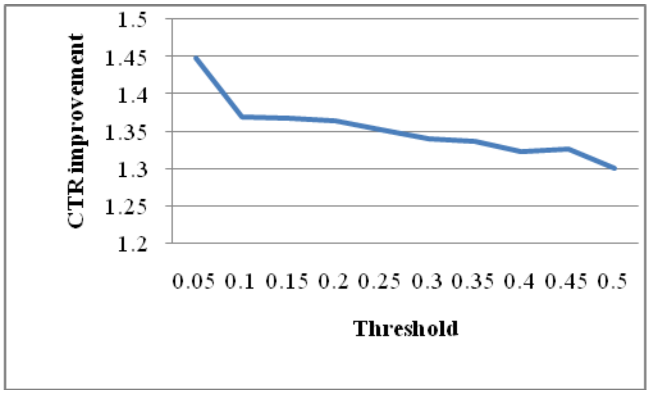

Figure 3. Change of CTR improvement with increasing threshold.

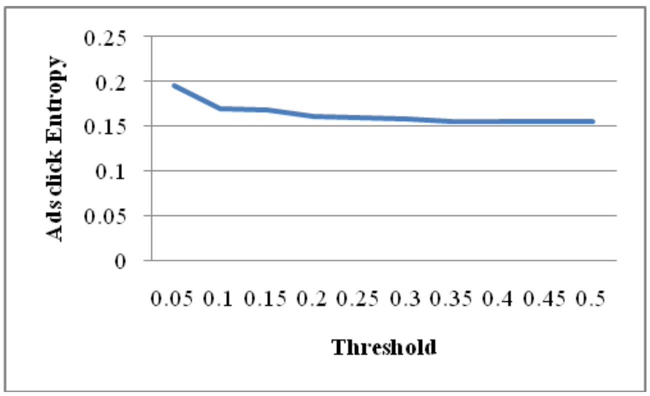

Figure 4. Change of ads click Entropy with increase threshold.

5. CONCLUSION AND FUTURE WORK

In this paper, we developed a novel semantic approach called PLSUS for BT. We compared the proposed PLSUS algorithm with two traditional clustering user segmentation approaches, CLUTO and k-Means. From the experimental results we can draw the conclusion that semantic approach PLSUS brings better improvements for BT in contrast to the traditional user clustering, especially in terms of CTR improvement.

In our future work, we will pay more attention to Latent Dirichlet (LDA). It has been noted that, LDA has better results in document classification than PLSA. Thus, we will study this model and attempt to apply it to user segmentation for verifying whether it has better performance for BT than PLSUS does. In addition, we will modify the EM algorithm to parallelize PLSUS. We believe it is helpful to further increase the algorithmic scalability and improve the efficiency.

6. REFERENCES

[1] D. Blei, A. Ng, and M. Jordan. Latent Dirichlet allocation. Journal of Machine Learning Research, 3(2003), 993–1022.

[2] T. Brants, F. Chen, and I. Tsochantaridis. Topic-based document segmentation with probabilistic latent semantic analysis. In Proceedings of CIKM '02 (Las Palmas, June 2002), ACM Press, 211-218.

[3] T. Brants and R. Stolle. Find similar documents in document collections. In Proceedings of LREC '02 (Spain, June 2002).

[4] A. Broder, M. Fontoura, V. Josifovski and L. Riedel. A semantic approach to contextual advertising. In Proceedings of SIGIR '07 (Amsterdam, July 2007), ACM Press, 559-566.

[5] D. Cohn and H. Chang. Learing to probabilistically identifying authoritative documents. In Proceedings of the ICML '00 (Stanford, June 2000), Morgan Kauffmann, 167-174.

[6] D. Cohn and T. Hofmann. The missing link: A probabilistic model of document content and hypertext connectivity. In Proceeding of NIPS '00 (Denver, November 2000), MIT Press.

[7] A. Das, M. Datar and A. Garg. Google news personalization: scalable online collaborative filtering. In Proceeding WWW '07 (Banff, May 2007), ACM Press, 271-280.

[8] S. Deerwester, S. Dumais, G. Furnas, T. Landauer, and R. Hashman. Indexing by latent semantic analysis. Journal of the American Society for Information Science, 41(1990), 391-407.

[9] D. C. Fain and J. O. Pedersen. Sponsored search: a brief history. In Bulletin of the American Society for Information Science and Technology, 2005.

[10] E. Gaussier, C. Goutte, K. Popat, and F. Chen. A hierarchical model for clustering and categorizing documents. In Advances in Information Retrieval – Proceedings of the 24th BCS-IRSG European Colloquium on IR Research (Glasgow, March 2002).

[11] M. Girolami and A. Kabán. On an equivalence between PLSI and LDA. In Proceeding SIGIR '03 (Toronto, July 2003), ACM Press, 433-434.

[12] A. Harpale and Y. Yang. Personalized active learning for collaborative filtering. In Proceeding of SIGIR '08 (Singapore, July 2008), ACM Press, 91-97.

[13] T. Hofmann. Probabilistic latent semantic analysis. In Proceedings of UAI '99 (Stockholm, July 1999), Morgan Kaufmann, 289-296.

[14] T. Hofmann. Probabilistic latent semantic indexing. In Proceedings of SIGIR '99 (Berkeley, August 1999), ACM Press, 50-57.

[15] T. Hofmann. Unsupervised learning by probabilistic latent semantic analysis. Machine Learning Journal, 42(2001), 177–196.

[16] X. Jin, Y. Zhou, and B. Mobasher. Web usage mining based on probabilistic latent semantic analysis. In Proceedings of KDD '04 (Seattle, August 2004), ACM Press, 197-205.

[17] Y. Kim, J. Chang, and B. Zhang. An empirical study on dimensionality optimization in text mining for linguistic knowledge acquisition. In Proceedings of KDD '03 (Seoul, April 2003), ACM Press, 111-116.

[18] G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. Information Processing and Management: an International Journal, 24 (1988), 513-523.

[19] J. Yan, N. Liu, G. Wang, W. Zhang, Y. Jiang and Z. Chen. How much the Behavioral Targeting can help online advertising? In Proceeding of WWW '09 (Madrid, April 2009), ACM Press, 261-270.

[20] Adlink, https://www.google.com/adsense/login/en_US/?gsessionid=Dc28hZShnCI

[21] Almond Net, http://www.almondnet.com/

[22] Blue Lithium, http://www.bluelithium.com/

[23] Burst, http://www.burstmedia.com/

[24] Double Click, http://www.doubleclick.com/products/dfa/index.aspx

[25] NebuAd, http://www.nebuad.com/

[26] Phorm, http://www.phorm.com/

[27] Revenue Science, http://www.revenuescience.com/advertisers/advertiser_solutions.asp

[28] Specificmeida, http://www.specificmedia.co.uk/

[29] TACODA, http://www.tacoda.com/

[30] Yahoo! Smart Ads, http://advertising.yahoo.com/marketing/smartads/