Vocabulary Expansion Technique for Advertisement Classification

Jin-Yong Jung, Jung-Hyun Lee, JongWoo Ha and SangKeun Lee
College of Information and Communications, Korea University
Seoul, 136-713, South Korea
[e-mail: {jyjung, jhbslpd, okcomputer, yalphy}@korea.ac.kr]
*Corresponding author: SangKeun Lee

Received October 16, 2011; revised December 22, 2011; revised February 8, 2012; accepted April 19, 2012; published May 25, 2012

Abstract

Contextual advertising is an important revenue source for major service providers on the Web. Ads classification is one of main tasks in contextual advertising, and it is used to retrieve semantically relevant ads with respect to the content of web pages. However, it is difficult for traditional text classification methods to achieve satisfactory performance in ads classification due to scarce term features in ads. In this paper, we propose a novel ads classification method that handles the lack of term features for classifying ads with short text. The proposed method utilizes a vocabulary expansion technique using semantic associations among terms learned from large-scale search query logs. The evaluation results show that our methodology achieves 4.0% ~ 9.7% improvements in terms of the hierarchical f-measure over the baseline classifiers without vocabulary expansion.

Keywords: Advertisement classification, vocabulary expansion, semantic association, query log, centroid classifier

This work was supported by Mid-career Researcher Program through NRF grant funded by the MEST (No. 2009-0077925)

http://dx.doi.org/10.3837/tiis.2012.05.007

1. Introduction

Contextual advertising is a form of online advertising, and has been an important revenue source for major service providers on the Web. Contextual advertising places textual ads within generic web pages, such as blog posts and online news articles. Textual ads consist of title, description, and url of the ad landing page, as illustrated in **Fig. 1**. In the prevalent pay-per-click pricing model, advertising revenue is gathered from advertisers when users click ads within a page, and the ad network that selects and places the ads in the page shares the revenue with the publisher of the page. Based on the observation that contextually relevant ads with the page have higher probabilities of being clicked by users than general or irrelevant ads [1][2], a lot of research on contextual advertising seeks to improve the relevance of retrieved ads [3][4][5][6].

The recent work [3] proposed a semantic approach to utilize category information to retrieve semantically relevant ads to a given web page. They first find the topics of the given web page and pre-collected ads by automatically classifying them into a common taxonomy of topics. Then, they ranks ads based on their topical relevance to the web page given the classification results. **Fig. 1** shows an example of the semantic approach to contextual advertising. In this example, the ad about a Chiang Mai resort is classified into the category "Recreation/Travel/Lodging/Hotels and Motels", and the given web page about a honeymoon in Chiang Mai is classified into the category "Recreation/Travel/Travelogues". Given these classification results, the system would place the ad in the web page because their topics are semantically relevant on the taxonomy.

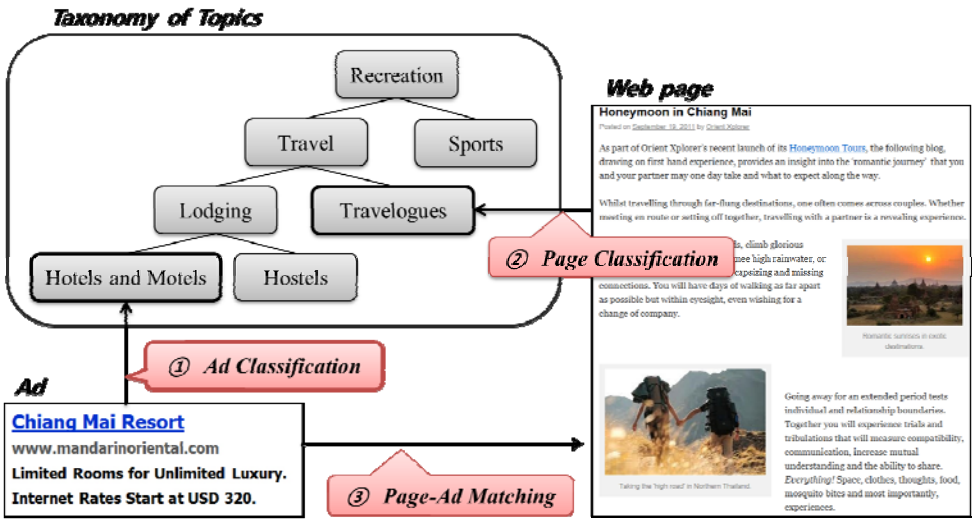

Fig. 1. Overview of the semantic approach to contextual advertising

In the semantic approach, the accurate classification of web pages and ads is crucial to the efficacy of topical relevance of page-ad matching. However, ads classification is relatively more difficult than web pages classification, due to the lack of useful term features in ads. We observed that it is difficult to obtain sufficient term features to classify ads since textual ads contain shorter text compared to general web pages. In addition, ads often contain terms

un-related to their topics. For example, advertisers emphasize the expected effect of their product (service) or price-related information to acquire users' attention, as shown in **Fig. 1**. This makes it difficult to obtain useful term features for classification; thereby leading to the poor classification performance (we will verify this in Section 4.). For instance, the ad in **Fig. 1** is classified into the irrelevant category "Computer/Internet" due to the price-related term "Internet".

Motivated by our observations, we address the problem of ads classification for contextual advertising. In particular, we handle with the scarcity of ads term features by using a novel vocabulary expansion technique. As shown in **Fig. 2**, our method consists of three tasks: 1) learning semantic associations from search query logs, 2) expanding ads with learned semantic associations, and 3) classifying with the original and expanded ads.

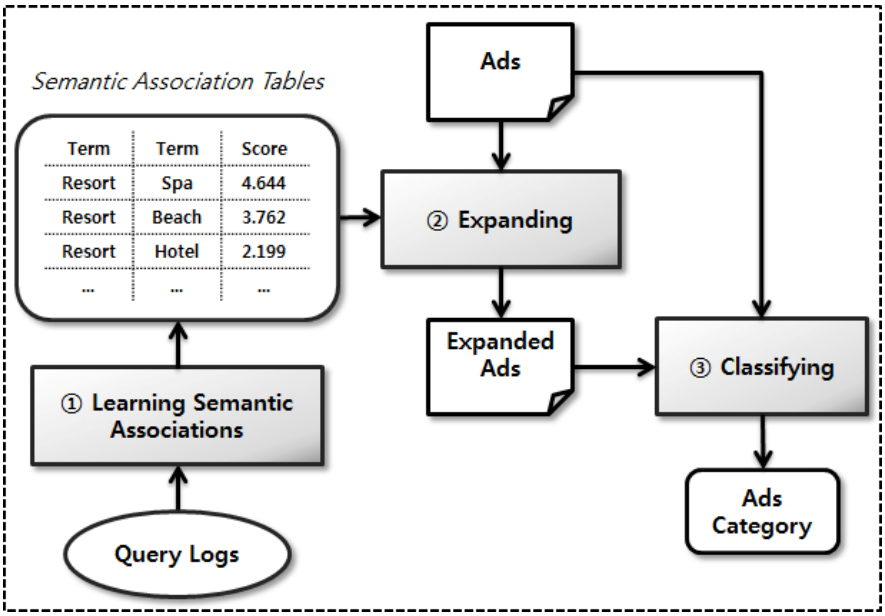

Fig. 2. Overview of proposed method

First, we learn semantic associations based on the co-occurrence calculated from *search query logs* (Section 3.1). We believe that search query logs are rich and valuable in learning semantic associations since query terms in each query are semantically associated to represent real users' information need. We utilize the search query logs to identify pairs of semantically associated terms, e.g., resort-spa, resort-beach, and resort-hotel. We then construct semantic association tables for each term with the learned semantic associations.

Second, we expand terms of an ad using the semantic association tables (Section 3.2). For example, if a given ad contains a term "resort", we augment term features in a way that the expanded ad contains "spa", "beach" and "hotel". For this task, we utilize two ad-specific features. One is a parameter τ . This is motivated by our observation that some ad terms are not useful to capture the topic of ad, e.g., "Internet" in the textual ad in **Fig. 1**. Thus, we utilize a parameter τ to filter irrelevant terms from expanded ad vocabulary. The other is a parameter *top-n*. This is used to select highly associated terms according to their semantic association with original ad terms. It also enables us to examine the number of expanded terms for effective ads classification.

Third, we develop a classification rule based on the centroid classifier by combining two classification results linearly, one obtained from the original ad and the other obtained from the expanded ad (Section 3.3). Note the expanded ad consists of the expanded terms only. Thus, the linear combination enables us to examine the efficacy of our vocabulary expansion technique in its ability to provide useful term features for the ads classification.

We conduct experiments to evaluate the proposed method using real-world textual ads collected from the Web. In the experiments, we also compare our method to other representative classifiers, k-Nearest Neighbor (kNN) classifier and hierarchical Support Vector Machine (SVM). The evaluation results show that our method achieves relatively 4.0% ~ 9.7% improvements in terms of the hierarchical f-measure over the baseline classifiers without vocabulary expansion.

In summary, the contributions of this paper include the following:

- We propose a novel ads classification method by using vocabulary expansion technique to overcome the scarcity of ads term features. To the best of our knowledge, this is the first study that addresses the problem of ads classification for contextual advertising.
- We utilize large-scale search query logs to learn useful semantic associations, and also exploit ad-specific features to effectively expand vocabulary.
- We conduct in-depth experiments to evaluate our methodology. The evaluation results show that our methodology improves ads classification performance as much as 4.0% ~ 9.7% over the baseline classifiers.

2. Related Work

Early work on contextual advertising was based on the word overlap within a page and ads to measure the relevance of ads [4][5]. Although these methods are effective in matching specific keywords, they lack the topical relevance of ads. To handle this problem, Broder at el. [3] proposed a semantic approach to contextual advertising using category information of a web page and ads to measure the topical relevance of ads to the page. The basic idea is to grasp the topics of a given web page and pre-collected ads by classifying them into a well-organized hierarchical taxonomy of topics. Given the classification results, their approach ranks ads based on their topical relevance to the page. Their evaluation result demonstrated that the category information could improve the relevance of ads significantly. Their work focused on the ad ranking methodology using category information and did not address ads classification as a research problem. In this paper, we focus on improving ads classification performance by applying the vocabulary expansion technique to handle the term scarcity of ads.

The vocabulary expansion technique is widely used for query expansion tasks for web search engines. The objective of query expansion is to provide sufficient information for an effective search of relevant documents. Query expansion is a traditional method, however, it is still utilized in many recent works mainly due to its effectiveness. For instance, it was utilized in a recent literature [7] to deal with the lack of term features in microblog search such as Twitter. Representative methods in query expansion include lexical-based methods [8][9], statistical-based methods [10][11][12], and query log-based methods [13][14].

The lexical-based methods utilize a manually created lexical thesaurus that includes information about synonyms and related terms. However, much cost is incurred in building the thesaurus, and the coverage of the thesaurus is limited. The statistical-based methods utilize the co-occurrence of terms in the document collection. Thus, the effectiveness of these

methods is heavily dependent on the quality of the collection. If the coverage of the collection is insufficient, the relationships between terms may not be well captured from the collection. Query log-based methods utilize the co-occurrence of terms in a set of queries and clicked documents recorded by web search activities of humans. Cui et al. [13] first proposed the query expansion method based on user interactions recorded in query logs. The main idea is to extract correlations between terms in queries and terms in clicked documents by analyzing query logs. Experimental results showed the query log-based query expansion method can improve the relevance of search results over the prior search method and other query expansion methods. The improvements of the query log-based methods are achieved from the advantage of real user judgments in query logs.

In this paper, we propose a mixed method of statistical approach and query log-based approach to expand insufficient ads vocabulary. In particular, our method learns semantic associations among terms in query logs using a representative statistical approach. In contrast with the previous query log-based method [13], we utilize query terms only for learning semantic associations to avoid invalid semantic associations which might be caused by clicked documents. In our preliminary investigation, we have observed that many invalid semantic associations are caused by clicked documents due to irrelevant and multi-topic documents. The invalid semantic associations degrade the classification performance of expanded ads. However, utilizing query terms only may lead to the poor coverage of extracted semantic associations. One possible solution is to additionally utilize clicked documents after filtering out irrelevant clicked documents with queries.

3. Proposed Method

3.1 Learning Semantic Associations

The first task of the proposed method is to learn semantic associations among individual terms from query logs. Query logs are collected by search engines and consist of queries issued by users and search results clicked by users in the real world. In general, terms in a query are semantically associated, since users specify keywords or terms in queries to describe their information needs. Thus, we expect that query logs will be good information sources to learn semantic associations among individual terms. For example, if two different terms are frequently co-occurred in many queries, their strong semantic association is indicated.

We use a representative association measure to learn semantic associations, *point-wise mutual information* (*PMI*) [15]. *PMI* is a measure of the relative entropy between the distributions of two terms, and it measures the extent to which two terms occur independently. Using *PMI* to learn semantic associations has been demonstrated being effective in many literatures. To expand original ad terms, we have also found that *PMI* is an effective measure to learn semantic associations from our preliminary experiments where *PMI* outperforms other semantic association measures such as χ^2.

Table 1. Incidence table of term t_i and term t_j

Case	**Queries that contain t_j**	**Queries that do not contain t_j**	**Total**
Queries that contain t_i	$n_{i,j}$	$n_i - n_{i,j}$	n_i
Queries that do not contain t_i	$n_j - n_{i,j}$	$N - n_i - (n_j - n_{i,j})$	$N - n_i$
Total	n_j	$N - n_j$	N

We first construct the incidence table of t_i and t_j, as shown in **Table 1**, to calculate *PMI* measure of term t_i and t_j. In this table, N is the number of queries in query logs, and n_i (n_j) is the number of queries that contain term t_i (t_j). $n_{i,j}$ is the number of queries that contain both t_i and t_j. Then, the *PMI* measure for two terms t_i and t_j is defined as follows:

$$PMI(t_i, t_j) = \log \frac{P(t_i, t_j)}{P(t_i)P(t_j)} = \log \frac{\frac{n_{i,j}}{N}}{\frac{n_i}{N} \times \frac{n_j}{N}}, \quad (1)$$

where $P(t_i)$ is the probability that term t_i occurs in a query, $P(t_j)$ is the probability that term t_j occurs in a query, and $P(t_i,t_j)$ is the probability that t_i and t_j occur in the same queries. If two terms t_i and t_j occur independently, $P(t_i,t_j) = P(t_i)P(t_j)$ and the *PMI* measure will be zero. If two terms tend to co-occur, $P(t_i,t_j)$ will be greater than $P(t_i)P(t_j)$ and the *PMI* measure will be higher.

We calculate an association score for each pair of terms in query logs using this association measure. We then generate a semantic association table for each term, where associated terms are indexed and ranked according to the association scores. **Table 2** illustrates a semantic association table of a term "resort". In addition, we apply two parameters in the semantic association tables to reduce noisy associations with low frequency in query logs. In particular, we exclude terms with lower query frequency than a parameter *min-f*, and terms with lower co-occurrence than a parameter *min-co*.

Table 2. Example of a semantic association table of the term "resort"

Rank	Term	Association score	Rank	Term	Association score
1	spa	4.644	6	lake	2.811
2	beach	3.762	7	golf	2.685
3	casino	3.634	8	island	2.656
4	grand	3.277	9	disney	2.605
5	bay	3.138	10	hotel	2.199

3.2 Expanding Ad

Given an original ad, the second task of the proposed method is to generate an expanded ad using the semantic association tables.

First, we represent the original ad as a vector of weighted occurrence frequency of individual terms based on the "Bag of Words" (BOW) model. The BOW model is the representative document representation model which has been widely used in traditional classification methods. Let A be the set of individual terms in the original ad: $A = \{a_1, a_2, a_3, ..., a_n\}$, then, the original ad is represented as a vector $\vec{a}$:

$$\vec{a} = [w(a_1), w(a_2), w(a_3), ..., w(a_n)], \quad (2)$$

where $w(a_k)$ is the weight of term a_k, and it is calculated by the standard *tf-idf* scheme:

$$w(a_k) = tf(a_k) \times \log(\frac{|D|}{df(a_k)}), \quad (3)$$

where $tf(a_k)$ is the frequency of term a_k within the original ad, D is a set of training documents and $df(a_k)$ is the document frequency of term a_k in D. We use the de-normalized term frequency in this weighting scheme because the lengths of ads are mostly consistent. In addition, the term frequency has more powerful ability to represent the topic especially in the short documents such as ads. We examined the performance of several normalization techniques for the term frequency in our preliminary experiments, and the de-normalized term frequency could achieve the best performance.

Second, we generate the expanded ad A' as a union of all sets of terms associated with each term in the original ad A:

$$A' = \{at \mid at \in S(a_1) \cup S(a_2) \cup S(a_3) \cup ... \cup S(a_n), at \notin A\}, \quad (4)$$

where $S(a_k)$ is the set of associated terms in the semantic association table of the term a_k. Note that the expanded ad A' does not include any original terms in the original ad A, that is, it only consists of associated terms extracted from query logs.

In this task, it is important to filter irrelevant terms with respect to the original context of ad, since irrelevant terms degrade the classification performance. We use two ad-specific features for this task. One is a parameter τ. We introduce this parameter by considering the characteristic of ad terms. As illustrated in **Fig. 1**, many ad terms emphasize the expected effect of their product (service) or price-related information, thus they are not useful features for ads classification. We conjecture that the irrelevant terms are expanded from less useful terms in the original ad and they should be filtered in our expanded ads. We use τ for this purpose. Specifically, if the normalized weight of $w(a_k)$ is less than τ, we filter $S(a_k)$ from the expanded ad A'. The other ad-specific feature is a parameter *top-n*. This parameter is used to select highly associated terms with original ad terms. We rank associated terms for term a_k on the basis of their association scores, and construct $S(a_k)$ with *top-n* associated terms.

Last, the expanded ad A' is also represented as a vector $\vec{a}'$:

$$\vec{a}' = [w(at_1), w(at_2), w(at_3), ..., w(at_n)], \quad (5)$$

where $w(at_i)$ is the weight of term at_i, and it is calculated by our proposed weighting scheme. We design the weighting scheme to let the weight of each associated term to be proportional to the weight of the original term and the association score with the original term. In the weighting scheme, a weight of an associated term at_i within A' is defined as the average of the original terms' weights multiplied by the normalized association scores:

$$w(at_i) = \frac{1}{|O(at_i)|} \sum_{a_k \in O(at_i)} w(a_k) \times NAS(a_k, at_i), \quad (6)$$

where $O(at_i)$ is the set of original terms associated with term at_i, $|O(at_i)|$ is the number of original terms associated with term at_i, $w(a_k)$ is the weight of term a_k in the original ad A, and $NAS(a_k, at_i)$ is a normalized association score for two terms a_k and at_i. The normalized association score is an association score for two terms a_k and at_i divided by a maximum among association scores of terms associated with term a_k:

$$NAS(a_k, at_i) = \frac{PMI(a_k, at_i)}{\max_{at \in S(a_k)}(PMI(a_k, at))}, \tag{7}$$

where $A(a_k)$ is the set of terms associated with term a_k, at is a term in set $A(a_k)$, and $PMI(a_k, at_i)$ is an association score for two terms a_k and at_i. **Fig. 3** illustrates how to calculate the weights of associated terms in the expanded ad. In this example, a circle represents a term, and a link between two circles represents an associated relation. The original ad A consists of two terms a_1 and a_2, and the expanded ad A' consists of seven terms from at_1 to at_7. The terms from at_1 to at_4 are associated with the original term a_1, and the terms from at_4 to at_7 are associated with the original term a_2. In the case of at_4, it is associated with two original terms a_1 and a_2, so the weight $w(at_4)$ is an average of weights propagated from two paths $a_1 \rightarrow at_4$ and $a_2 \rightarrow at_4$.

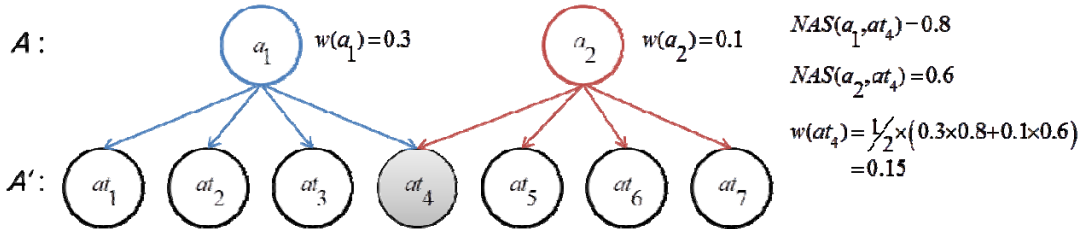

Fig. 3. Example for calculating the weights of associated terms in the expanded ad

3.3 Classification

The third and last task of the proposed method is to classify the original ad and the expanded ad into a set of pre-defined categories. In this paper, we aim to classify ads into a large-scale set of categories such as Open Directory Project (ODP) [16], since it is more effective for improving the relevance of ads to provide detail category information for ads in the contextual advertising [3]. There are many classification algorithms such as centroid classifier, k-Nearest Neighbor (kNN) classifier, Naïve Bayesian classifier, and SVM to train classifiers for such a large-scale set of categories. Although SVM is known to have the best classification performance, it is not proper for the ad classification task because it has low efficiency to train and test the classifier for large-scale categories. Instead, we use the centroid classifier [17] based on vector space model [18], since the recent work [3] in contextual advertising adapted it for the ad classification task and the another work [19] shows that the centroid classifier is effective and efficient for a large-scale set of categories. In Section 4.3, we also give the experimental results from other classifiers with expanded ads.

The classification rule of the centroid classifier is to classify an ad in accordance with the category region to which it belongs. The category region for each category is computed from its training documents, and it is denoted as *centroid*. The centroid $\vec{c}_i$ of a category c_i is defined as follows:

$$\vec{c}_i = \frac{1}{|D_i|} \sum_{d_j \in D_i} \vec{d}_j, \tag{8}$$

where D_i is a set of training documents for the category c_i, and $\vec{d}_j$ is a normalized term vector of a training document d_j in D_i. The centroid $\vec{c}_i$ is an average vector of all training documents in the category c_i.

The classifier computes distances between the ad and all centroids, and then assigns the ad to category c whose centroid $\vec{c}$ is the closest to it. We use the cosine similarity as the distance function. We define the final category assignment criterion to combine classification results of the original ad a and expanded ad a', as follows:

$$\arg\max_{c_i} ((1-\alpha)\times\cos(\vec{a},\vec{c}_i)+\alpha\times\cos(\vec{a}',\vec{c}_i)), \quad (9)$$

where $\vec{c}_i$ is the centroid of category c_i, and α is the weight for linear combination between the original ad's classification results and the expanded ad's classification results. In the experiment, we evaluate the classification performance of the proposed method on a variety of α values to find the best linear combination.

4. Evaluation

4.1 Data and Performance Measures

To evaluate the proposed method, we first collect a training set for classifiers. The pioneering work [3] used a training set which consists of a commercial taxonomy of around 6,000 nodes manually built by Yahoo! US and bid phrases manually classified into the taxonomy. In contrast with their approach, we collect a training set from the Open Directory Project (ODP) [16] to reduce an excessive amount of human effort for manually constructing the training set. ODP is a large-scale and tree-structured web directory. ODP is widely used for various classification tasks, including web page classification [20][21][22], query classification [23], and personalized search [24], due to its large-scale, high-quality, and availability. ODP currently consists of 771,762 categories and 4,375,354 web pages classified into categories by human experts. Categories under "Top/Regional/" and "Top/World/" are removed in our training set, since those categories duplicate each other and are not written in English. We only targeted English in this evaluation, although our method can be applied to other languages. Thus, the training set consists of 182,042 categories and 1,347,567 web pages.

Second, we construct a test set that consists of textual ads each of which has ODP categories as labels. We collect 89,637 advertising keywords by submitting category names in the training set to an ad keyword suggestion tool[1] of Google AdWords. We then collect 156,235 distinct ads from an ad search[2], using these keywords as queries. Next, we label categories for each ad. We search categories that contain web page with the same ad url from the training set, and assign the labels as the categories for each ad and its url. Thus, we obtain a test set consisting of 46,240 distinct ads with category labels. We also construct a validation set for parameter setting with randomly selected 1,156 ads.

Third, we use the query log included in MSN search log data to learn semantic associations. The query log was released by Microsoft Live Labs in 2006. It spans 30 days from 05/01/2006 to 05/30/2006. The total number of queries is 8,831,280 and the number of distinct queries is

[1] http://adwords.google.com/select/keywordtoolexternal
[2] http://www.google.com/sponsoredlinks

3,874,994. We learn semantic associations for 712,068 distinct terms from the query log. This covers over 80.2% of ad terms in the test set.

We measure the classification performance of the proposed method on the large-scale category hierarchy in terms of hierarchical precision (hP), hierarchical recall (hR), and hierarchical f-measure (hF) proposed by Kiritchenko et al. [25]. Each of three metrics is defined as follows:

$$hP = \frac{\sum_i |P_i \cap T_i|}{\sum_i |P_i|}, hR = \frac{\sum_i |P_i \cap T_i|}{\sum_i |T_i|}, hF = \frac{2 \times hP \times hR}{hP + hR}, \quad (10)$$

where P_i is the set with the most specific category predicted for test ad i and all its ancestor categories, and T_i is the set with the true most specific categories of test ad i and all its ancestor. These metrics are known to be effective to compare the performance of different algorithms on the hierarchical classification task [26].

4.2 Parameter Setting

In this section, we present experimental results on the validation set for parameter setting. We use several parameters in the proposed method. First, parameter *min-f* and *min-co* are used to remove noisy information in the task to learn semantic associations from the query log. We empirically set *min-f* to 100 and *min-co* to 10 to obtain co-occurrence information of terms that frequently occur in the query log.

Second, parameter τ is used to select ad terms that will be expanded according to the importance of terms. **Table 3** shows the ad classification performance on different τ values from 0 to 0.9. Other parameters are fixed in this experiment. We only present experimental results for the combination of other parameters with best results, although we conducted experiments for all combinations of the parameters. As shown in **Table 3**, we can achieve the best performance when parameter τ is 0.2.

Lastly, parameter *top-n* is used to control the size of the expanded ad. **Table 4** shows the ad classification performance on different *top-n* values from 10 to 110. As shown in this table, the performance rapidly increases when parameter *top-n* increases from 10 to 40, and the performance gradually increases when parameter *top-n* increases from 50 to 100. We can achieve the best performance when parameter *top-n* is 100.

Table 3. Performance with different τ on the validation set (*top-n*=100, α =0.6)

τ	**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
hP	0.288	0.289	**0.290**	0.284	0.277	0.266	0.244	0.230	0.221	0.204
hR	0.242	0.243	**0.245**	0.241	0.235	0.229	0.211	0.203	0.199	0.184
hF	0.263	0.264	**0.265**	0.260	0.254	0.246	0.227	0.216	0.209	0.194

Table 4. Performance with different *top-n* on the validation set (τ =0.2, α =0.6)

top-n	10	20	30	40	50	60	70	80	90	100	110
hP	0.228	0.258	0.268	0.277	0.280	0.280	0.282	0.285	0.288	**0.290**	0.289
hR	0.188	0.213	0.221	0.230	0.232	0.233	0.235	0.238	0.242	**0.245**	0.244
hF	0.206	0.233	0.242	0.252	0.254	0.254	0.256	0.259	0.263	**0.265**	**0.265**

4.3 Performance On Different Alpha

In this section, we evaluate our proposed method based centroid classifier with different α values on the test set. **Table 5** presents the classification performance with various α values from 0 to 1. α is a parameter that controls the relative weight assigned to the classification score obtained from the expanded ad. Thus, we have the performance using the expanded ad only at α =1. In contrast, we have performance using the baseline without vocabulary expansion at α =0. Note the expanded ads consist of the expanded terms only, and basically they do not include terms in the original ads. Thus, this evaluation is also designed to examine the impact of expanded terms with various α values. We expect that we can find the best performance in the middle of the linear combination.

As shown in **Table 5**, the performance of the proposed method based centroid classifier increases when weight α increases from 0 to 0.5. It demonstrates that the original ads have few term features, so it requires more useful terms for the classification. It also demonstrates that the expanded terms in our method are effective to enrich term features of the original ads. The performance at α =1 signifies the importance of terms presented in the original ads. Since the expanded ads do not include terms in the original ads, the classification result obtained from only the expanded terms is poor than the classification result obtained from the original ad terms. The hF performance of the baseline is 0.238, and the hF performance of the best combination is 0.261 at α =0.5 and α =0.6. Thus, the proposed method improves the performance up to 9.7% compared to the baseline. This confirms that our vocabulary expanding technique is an effective method to handle the scarcity of the term features for classifying ads. The result in **Table 5** also signifies the importance of setting the effective parameter as the performance at α larger than 0.8 is poorer than for the baseline.

Table 5. Performance with different α on the test set(τ =0.2, *top-n*=100)

α	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
hP	0.271	0.275	0.279	0.284	0.289	**0.290**	0.289	0.277	0.250	0.219	0.193
hR	0.212	0.216	0.222	0.228	0.233	0.236	**0.238**	0.232	0.210	0.182	0.159
hF	0.238	0.242	0.247	0.253	0.258	**0.261**	**0.261**	0.252	0.228	0.199	0.174

4.4 Performance On Different Classification Methods

We evaluate the effectiveness of our ad vocabulary expansion method on different classification methods. In this experiment, we use two other representative classifiers,

k-Nearest Neighbor (kNN) classifier[3] and hierarchical Support Vector Machine (SVM)[4] [27]. We compare the performance of each classifier using both expanded ads and original ads to the base case with only original ads.

Table 6 presents the evaluation results of different classification methods, centroid classifier, i.e., our proposed method, kNN classifier and hierarchical SVM. In this table, classifiers using both expanded ads and original ads are denoted as Centroid-E, kNN-E, and Hierarchical SVM-E, respectively. As shown in this table, all classifiers using both expanded ads and original ads achieve about 4.0~9.7% improvements over the base classifiers with only original ads in terms of the hF measure. These results demonstrate that our ad vocabulary expanding method is also effective on different classification methods. Especially, our proposed method, Centroid-E, achieves the largest improvements on the hF measure and the hR measure. Also, the hF performance of Centroid-E is similar to the performance of Hierarchical SVM-E which is known to have the best classification performance in many literatures. The hierarchical SVM classifiers have the highest hierarchical precision performance, however, the lowest hierarchical recall performance, since they predicted most ads into categories at upper levels in the category hierarchy. The performance of SVM heavily depends on the number of positive examples. Thus, the hierarchical SVM classifiers, which are trained from our training set where many categories at low levels have few positive examples, have the tendency to stop the classification at upper levels in the category hierarchy. These results also demonstrate that our proposed method based on the centroid classifier is effective in classifying ads into the large-scale category hierarchy.

Table 6. Evaluation results of different classification methods

Classifier	Centroid	Centroid-E	kNN	kNN-E	Hierarchical SVM	Hierarchical SVM-E
hP	0.271	0.289 (+6.6%)	0.259	0.267 (+2.8%)	0.455	**0.491 (+8.5%)**
hR	0.212	**0.238 (+12.2%)**	0.209	0.220 (+5.0%)	0.161	0.175 (+8.7%)
hF	0.238	**0.261 (+9.7%)**	0.232	0.241 (+4.0%)	0.238	0.259 (+8.7%)

5. Conclusion

In this paper, we developed a novel ads classification method for contextual advertising. We first learned semantic associations by calculating PMI from search query logs to handle the scarcity of term features. We then applied the vocabulary expansion technique to obtain expanded ads using semantic association tables. Finally, we combined the two classification scores obtained from the original ad and expanded ad using the centroid classifier. We confirmed PMI calculated from search query logs is an effective measure of semantic

[3] We set k to 13 empirically after inspecting performance results on the validation set.
[4] We implemented the hierarchical SVM using SVMLightLib (http://mihagrcar.org/svmlightlib/). We trained linear SVM classifier per each category in the hierarchy using the "siblings" policy [26] for selecting positive and negative examples. Then, we used the top-down approach with the non-mandatory leaf-node prediction (NMLNP) policy for classifying test ads.

association from the evaluation. The evaluation results demonstrated our method significantly improves ads classification performance.

References

[1] P., Chatterjee, D. L. Hoffman,and P. T. Novak, "Modeling the clickstream: Implications for web-based advertising efforts," *Marketing Science*, vol.22, no.4, pp.520-541, 2003. Article (CrossRef Link)

[2] C. Wang, P. Zhang, R. Chol and M. D'eredita, "Understanding Consumers Attitude Toward Advertising," in *Proc. of 8th Americas Conf. on Information Systems*, pp.1143-1148, 2002. Article (CrossRef Link)

[3] A. Broder, M. Fontoura, V. Josifovski, and L. A. Riedel, "Semantic approach to contextual advertising," in *Proc. of 30th Int. ACM SIGIR Conf. on Research and development in information retrieval*, pp.559-566, 2007. Article (CrossRef Link)

[4] W.-T. Yih, J. Goodman and V. R. Carvalho, "Finding advertising keywords on web pages," in *Proc. of 15th International Conference on World Wide Web*, pp.213-222, 2006. Article (CrossRef Link)

[5] B. A. Ribeiro-Neto, M. Cristo, P. B. Golgher and E. S. De Moura, "Impedance coupling in content-targeted advertising," in *Proc. of 28th International ACM SIGIR Conference on Research and development in information retrieval*, pp.496-503, 2005. Article (CrossRef Link)

[6] V. Murdock, M. Ciaramita and V. A. Plachouras, "Noisy-channel approach to contextual advertising," in *Proc. of 1st Int. Workshop on Data Mining and Audience Intelligence for Advertising*, pp.21-27, 2007. Article (CrossRef Link)

[7] M. Karam, T. Manos, d. R. Marten and W. Wouter, "Incorporating query expansion and quality indicators in searching microblog posts," in *Proc. of 33rd Eur. Conference on Advances in Information Retrieval*, pp.362-367, 2011. Article (CrossRef Link)

[8] E. M. Voorhees, "Query expansion using Lexical-Semantic relations," in *Proc. of 17th Int. ACM SIGIR Conerence on Research and development in information retrieval*, pp.61-69, 1994. Article (CrossRef Link)

[9] O.-W. Kwon, and M.-C. Kim and K.-S. Choi, "Query expansion using domain adapted thesaurus in an extended boolean model," in *Proc. of 3rd ACM Int. Conf. on Information and Knowledge Management*, pp.140-146, 1994. Article (CrossRef Link)

[10] Y. Qiu, and H. P. Frei, "Concept based query expansion," in *Proc. of 16th International ACM SIGIR Conerence on Research and development in information retrieval*, pp.160-169, 1993. Article (CrossRef Link)

[11] J. Xu and W. B. Croft, "Improving the effectiveness of information retrieval with local context analysis," *ACM Transactions on Information Systems*, vol.18, no.1, pp.79-112, 2000. Article (CrossRef Link)

[12] J. Bai, D. Song, P. Bruza, J.-Y. Nie and G. Cao, "Query expansion using term relationships in language models for information retrieval," in *Proc. of 14th ACM Int. Conference on Information and Knowledge Management*, pp.688-695, 2005. Article (CrossRef Link)

[13] H. Cui, J.-R. Wen, J.-Y. Nie and W.-Y. Ma, "Query expansion by mining user logs," *IEEE Transactions on Knowledge and Data Engineering*, vol.15, no.4, pp.829-839, 2003. Article (CrossRef Link)

[14] B. Billerbeck, F. Scholer, H. E. Williams and J. Zobel, "Query expansion using associated queries," in *Proc. of 12th ACM Int. Conf. on Information and Knowledge Management*, pp.2-9, 2003. Article (CrossRef Link)

[15] C. J. van Rijsbergen, "Information retrieval," London: Butter-worths, 1979.

[16] The Open Directory Project, http://www.dmoz.org/.

[17] E.-H. S. Han and G. Karypis, "Centroid-based document classification: Analysis and experimental results," in *Proc. of Eur. Conference on Principles and Practice of Knowledge Discovery in Databases*, pp.424-431, 2000. Article (CrossRef Link)

[18] G. Salton, A. Wong and C. Yang, "A vector space model for automatic indexing," *Communication ACM*, vol.18, no.11, pp.517-526, 1975. Article (CrossRef Link)
[19] E. Gabrilovich and S. Markovitch, "Feature generation for text categorization using world knowledge," in *Proc. of 19th International Joint Conf. on Artificial Intelligence*, pp.1048-1053, 2005. Article (CrossRef Link)
[20] P. N. Bennett and N. Nguyen, "Refined experts: Improving classification in large taxonomies," in *Proc. of 32nd International ACM SIGIR Conference on Research and development in information retrieval*, pp.11–18, 2009. Article (CrossRef Link)
[21] J.-J. LEE, J.-H LEE, J. HA and S. LEE, "Novel web page classification techniques in contextual advertising," in *Proc. of 7th International Workshop on Web Information and Data Management*, pp.39–47, 2009. Article (CrossRef Link)
[22] G. R. Xue, D. Xing, Q. Yang and Y. Yu, "Deep classification in large-scale text hierarchies," in *Proc. of 31st International ACM SIGIR Conference on Research and development in information retrieval*, pp.619-626, 2008. Article (CrossRef Link)
[23] P. N. Bennett, K. Svore, and S. T. Dumais, "Classification-enhanced ranking," in *Proc. of 19th International Conference on World Wide Web*, pp.111–120, 2010. Article (CrossRef Link)
[24] P. A. Chirita, W. Nejdl, R. Paiu and C. Kohlschutter, "Using ODP metadata to personalized search," in *Proc. of 28th International ACM SIGIR Conference on Research and development in information retrieval*, pp.178–185, 2005. Article (CrossRef Link)
[25] S. Kiritchenko, S. Matwin and AF. Famili, "Functional annotation of genes using hierarchical text categorization," in *Proc. BioLINK SIG Meeting on Text Data Mining at ISMB'05*, 2005. Article (CrossRef Link)
[26] S. Carlos and F, Alex, "A survey of hierarchical classification across different application domains," *Data Mining and Knowledge Discovery*, vol.22, no.1, pp.31-72, 2011. Article (CrossRef Link)
[27] T.-Y. Liu, Y. Yang, H. Wan, H.-J. Zeng, Z. Chen and W.-Y Ma, "Support vector machines classification with a very large-scale taxonomy," *SIGKDD Explor. Newsl.*, vol.7, no.1, pp.36-43, 2005. Article (CrossRef Link)

Jin-Yong Jung received the B.S. degree in forestry as the major and in computer science and engineering as the minor from Dongguk University in 1993 and the M.S. degree in computer science and engineering from Korea University in 2005. He is a Ph.D. candidate in the College of Information and Communications, Korea University. His research interests include data management in contextual advertising.

Jung-Hyun Lee received the B.S. and M.S. degrees in computer science and engineering from Korea University in 2009 and 2011, respectively. He is a Ph.D. student in the College of Information and Communications, Korea University. His research interests include text classification in large-scale topic hierarchies and data management in contextual advertising and personalized services.

JongWoo Ha received the B.S. and M.S. degrees in computer science and engineering from Korea University in 2007 and 2009, respectively. He is a Ph.D. candidate in the College of Information and Communications, Korea University. His research interests include data management in contextual advertising and personalized services.

SangKeun Lee received the B.S., M.S., and Ph.D. degrees in computer science and engineering from Korea University in 1994, 1996, and 1999, respectively. He was a recipient of the Japan Society for the Promotion of Science (JSPS) Postdoctoral Fellowship in 2000. Since 2003, he has been an assistant/associate/full professor in the College of Information and Communications, Korea University. His research interests include data management in mobile/pervasive computing systems, location-based information systems, XML databases, and data management in mobile ad hoc networks.