

# Query Rewriting using Active Learning for Sponsored Search

Wei Vivian Zhang, Xiaofei He, Benjamin Rey and Rosie Jones  
Yahoo!  
3333 Empire Avenue  
Burbank, CA, 91504, USA  
{zhangv,hex}@yahoo-inc.com

## ABSTRACT

Sponsored search is a major revenue source for search companies. Web searchers can issue any queries, while advertisement keywords are limited. Query rewriting technique effectively matches user queries with relevant advertisement keywords, thus increases the amount of web advertisements available. The match relevance is critical for clicks. In this study, we aim to improve query rewriting relevance. For this purpose, we use an active learning algorithm called Transductive Experimental Design to select the most informative samples to train the query rewriting relevance model. Experiments show that this approach improves model accuracy and rewriting relevance.

## Categories and Subject Descriptors

H.3.3 [Information Search and Retrieval]: Query formulation

## General Terms

Algorithm, Design, Experimentation

## Keywords

Query rewriting, active learning, sponsored search

## 1. INTRODUCTION

Sponsored search marketing is an important business for on-line advertising. On one hand, advertisers choose keywords to describe their advertisements. The keyword set is within a limit size. On the other hand, web searchers issue huge amount of queries per day. The search query set could be infinite large. One challenge for sponsored search is to match user queries with advertisement keywords.

Query rewriting is an effective approach to approximately match between user queries and advertisement keywords. The query rewriting system reformulates user queries into different forms (i.e. rewrites), and these new forms can be used as advertisement keywords to fetch sponsored results. In previous work, machine learning models were developed for rewrite relevance ranking. These models were trained using human editorial relevance judgments. The hand-labeling process is expensive and thus limited.

In our study, we use a Transductive Experimental Design [2] to select the most informative samples for labeling

and training. Our results show that this approach improves model performance with equal amount of labeled samples, compared with the previous query rewriting method [1]. In addition, no labels are required for selecting the most informative training examples, which is different from many other active learning methods. This property is especially important for real application, where labeling and modeling are usually performed separately due to the consideration of resource efficiency. The main contribution of this paper is to apply active learning technique to query rewriting.

## 2. QUERY REWRITING FOR SPONSORED SEARCH

Jones et al [1] proposed a query rewriting system for sponsored search. First, the candidate set is generated by mining query logs and extracting frequent sequential queries issued by web searchers. This is the source of related queries or phrases. Second, a machine-learning model is developed to rank rewrite candidates in terms of their relevance to user original query. The most relevant rewritten queries are then used as keywords to fetch advertisements shown.

In order to train the relevance model, Jones et al [1] randomly sampled (original query, rewritten query) pairs. These pairs were hand-labeled by an editorial team using the four-point scale of relevance judgments given in Table 1.

## 3. QUERY REWRITING USING ACTIVE LEARNING

### 3.1 Feature Generation

For each (original query, rewritten query) pair, we generate the following features and use them as model predictor variables: 1. **Syntactic features**: these features include edit distance, token overlap and character prefix overlap between the original query and its rewrite; 2. **Statistical rewrite frequency features**: the frequency and probability with which the original query is reformulated into its rewrite. These features are computed by mining query logs; 3. **Web occurrence feature**: the frequency with which the original query and its rewrite appear on same web pages; 4. **Web search feature**: the frequency with which the original query and its rewrite share common URLs as their search results; 5. **Advertiser bid feature**: the frequency with which the original query and its rewrite are bid by the same advertisers; 6. **Location feature**: whether there is location change between the original query and its rewrite. Each training pairs is labeled using the guidelines shown in Table 1.

Score	Type	Definition	Example
1	Precise Match	A near-certain match.	automotive insurance - automobile insurance
2	Approximate Match	A probable, but inexact match with user intent.	hybrid car - toyota prius
3	Marginal Match	A distant, but plausible match to a related topic.	ibm thinkpad - laptop bag
4	Mismatch	A clear mismatch.	time magazine - time and date magazine

Table 1: Editorial Scoring System for Query Rewrites

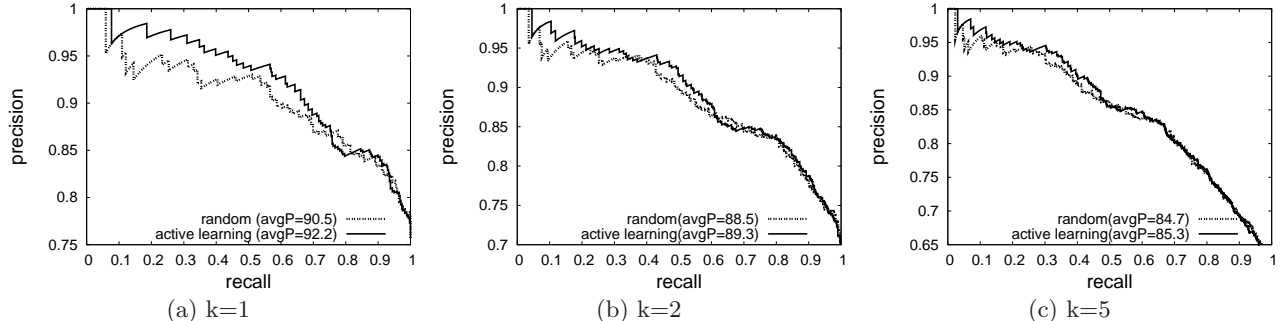


Figure 1: Precision-recall for top  $k$  ( $=1,2,5$ ) best rewrites.

### 3.2 Selecting the Most Informative Samples

In this section, we apply an active learning approach called Transductive Experimental Design (TED, [2]) to select the most informative (query, rewrite) pairs. The reason we choose TED is that it explicitly considers both measured and unmeasured samples. Since the relevance of a query pair  $\{q, q'\}$  is characterized by the feature vector  $\mathbf{x}$ , we use  $\mathbf{x}_i$  to denote the query pair  $\{q_i, q'_i\}$  hereafter. Let  $\mathcal{X}$  denote the set of all feature vectors, i.e.  $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$ .

Consider a linear regression model

$$y = \mathbf{w}^T \mathbf{x} + \epsilon \quad (1)$$

where  $y$  is the *observation*,  $\mathbf{x}$  is the *predictor variable*,  $\mathbf{w}$  is the *weight vector* and  $\epsilon$  is an unknown error with zero mean and  $\sigma^2$  variance. We define  $f(\mathbf{x}) = \mathbf{w}^T \mathbf{x}$  to be the learner's output (relevance score) given input  $\mathbf{x}$  and the weight vector  $\mathbf{w}$ . Given a subset  $\mathcal{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_k\} \subset \mathcal{X}$  ( $k < m$ ), let  $y_i$  be the relevance score associated with  $\mathbf{z}_i$ . The maximum likelihood estimate for the weight vector,  $\hat{\mathbf{w}}$ , is that which minimizes the sum squared error:

$$\hat{\mathbf{w}}_{\mathcal{Z}} = \arg \min_{\mathbf{w}} \sum_{i=1}^k (\mathbf{w}^T \mathbf{z}_i - y_i)^2 \quad (2)$$

We define  $Z = (\mathbf{z}_1, \dots, \mathbf{z}_k)$  and  $\mathbf{y} = (y_1, \dots, y_k)$ , thus  $\hat{\mathbf{w}}_{\mathcal{Z}}$  is given by

$$\hat{\mathbf{w}}_{\mathcal{Z}} = (ZZ^T)^{-1} Z^T \mathbf{y}$$

It is easy to check that  $E(\hat{\mathbf{w}}_{\mathcal{Z}}) = 0$ . Therefore,  $\hat{\mathbf{w}}$  is an unbiased estimate. For any new  $\mathbf{x}$ , let  $\hat{y} = \hat{\mathbf{w}}^T \mathbf{x}$  be its predicted observation. The expected square prediction error can be written as follows:

$$E(y - \hat{y})^2 = \sigma^2 \mathbf{x}^T (ZZ^T)^{-1} \mathbf{x}$$

Clearly the expected square prediction error depends on the independent variable  $\mathbf{x}$ , therefore the average expected square predictive error over the complete data set  $\mathcal{Z}$  is

$$\frac{1}{k} \sum_{i=1}^k E(y_i - \hat{\mathbf{w}}^T \mathbf{z}_i)^2 = \sigma^2 + \frac{1}{k} \sigma^2 \text{Tr}(X^T (ZZ^T)^{-1} X) \quad (3)$$

In order to minimize the average expected square predictive error, one should find a subset  $Z$  which minimizes Eqn. (3).

## 4. EXPERIMENTAL RESULTS

We compare our approach with random sampling for learning a linear query rewriting relevance model. We apply both random sampling and TED to select 1500 samples from a pool of 5000 examples for labeling and training, and use 1500 held-out examples for testing. We evaluate on a classification task. The positive class includes label score 1 and 2 (i.e. precise and approximate match as mentioned in Table 1). The negative class includes score 3 and 4 (i.e. marginal match and mismatch).

Figure 1 shows the precision-recall curves for top  $k$  ( $=1, 2, 5$ ) best query rewrites. As can be seen, our approach consistently outperforms random sampling. Especially, our approach improves the prediction accuracy when only the best rewrite is considered ( $k=1$ ). In this case, we achieve 1.7% improvement in average precision, or 18% error reduction.

## 5. CONCLUSIONS

We used the active learning algorithm to select the most informative samples to train the relevance model for query rewriting. Our approach improves the model prediction accuracy and query rewriting quality. In practice, our approach can be applied in sponsored search to improve the relevance between user queries and matched advertisements keywords. It can also reduce labor-intensive labeling task by eliminating non-informative training samples, and thus improve production efficiency.

## 6. REFERENCES

- [1] R. Jones, B. Rey, O. Madani, and W. Greiner. Generating query substitutions. In *Proceedings of WWW*, Edinburgh, Scotland, 2006.
- [2] K. Yu, J. Bi, and V. Tresp. Active learning via transductive experimental design. In *Proceedings of the 23rd International Conference on Machine Learning*, 2006.