Automatically Generating Labels Based on Unified Click Model

Guichun Hua, Min Zhang, Yiqun Liu, Shaoping Ma, Liyun Ru
State Key Laboratory of Intelligent Technology and Systems
Tsinghua National Laboratory for Information Science and Technology
Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
huaguichun@gmail.com, {z-m, yiqunliu, msp}@tsinghua.edu.cn, lyru@vip.sohu.com

ABSTRACT

Ground truth labels are one of the most important parts in many test collections for information retrieval. Each label, depicting the relevance between a query-document pair, is usually judged by a human, and this process is time-consuming and labor-intensive. Automatically Generating labels from click-through data has attracted increasing attention. In this paper, we propose a Unified Click Model to predict the multi-level labels, which aims at comprehensively considering the advantages of the Position Models and Cascade Models. Experiments show that the proposed click model outperforms the existing click models in predicting the multi-level labels, and could replace the labels judged by humans for test collections.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Retrieval Model

General Terms

Algorithms, Performance, Experimentation.

Keywords

Click Model, Learning to Rank, Ranking SVM

1. INTRODUCTION

Many test collections have been builded as benchmark datasets for researchers comparing the performance of the models, such as $TREC$ test collections and $LETOR$ [6]. Each label in these test collections, which dipicts the relevance between a query-document pair, is usually judged by a human, and this process is very time-consuming and labor-intensive. Many models, automatically generating labels from click-through data, have been proposed which could be classified into two categories: Position Models and Cascade Models.

Position Models, such as [9], assume that the users' examining and clicking a URL in the returning result depends only on the position of this URL. While Cascade Models, such as [3], consider that the URLs ranking above the clicked URL have effects on the clicked one. Cascade Models are more useful when fewer URLs are clicked; otherwise Position Models perform better. The users have diverse intention requirements and the queries submitted by users could be navigational, informational and transactional [1, 7]. For navigational queries, users usually click a fewer URLs, otherwise for informational and transactional queries. In this circumstance, it is not reasonable to use a general click model to predict the relevance between a query-document pair.

The main contributions of this paper are that we propose a Unified Click Model to predict the label between a query-document pair comprehensively considering the advantages of Position Models and Cascade Models. The experimental results show that the labels predicted by the Unified Click Model could replace the labels judged by humans and performs better than the labels predicted by the Position Model or Cascade Model alone.

WWW 2011, March 28–April 1, 2011, Hyderabad, India.
ACM 978-1-4503-0637-9/11/03.

2. UNIFIED CLICK MODEL

The Unified Click Model contains two parts: Query Intention Identification and Click Models. The Query Intention Identification predicts the query intention, and for each type of queries, different Click Models are chosen to predict the relevance between a query-document pair. The Query Intention Identification is the bridge to connect different kinds of Click Models.

2.1 Query Intention Identification

The decision tree $C4.5$ algorithm is chosen as the query intention classifier with two useful features nCS(n Clicks Satisfied) and nRS(Top n Results Satisfied) which are proved to be very efficient in [8]. The nCS and nRS are defined as below.

$$nCS(q) = \frac{\#(Session\ of\ q\ less\ than\ n\ clicks)}{\#(Session\ of\ q)}$$

$$nRS(q) = \frac{\#(Session\ of\ q\ clicks\ only\ on\ top\ n\ results)}{\#(Session\ of\ q)}$$

$\#(Session\ of\ q\ less\ than\ n\ clicks)$ means the number of sessions, involving the query q, in which every one has the number of clicks less than n. $\#(Session\ of\ q)$ means the number of sessions involving the query q. And $\#(Session\ of\ q\ clicks\ only\ on\ top\ n\ results)$ means the number of sessions, involving the query q, in which every one has clicks only on top n results.

2.2 Click Models

The Position Model [9], referred to as $COEC$ (clicks over expected clicks), and the Cascade Model [2], referred as DBN (dynamic bayesian network click model), are chosen as the click models in this paper.

2.3 Using Unified Click Model to Predict Multi-Level Labels

In the application of the Unified Click Model, we use the C4.5 algorithm to learn the thresholds to map the relevance predicted by click models into discrete label values. In this way, the Unified Click Model used to predict multi-level labels is a decision tree with features of nCS, nRS, $COEC$, DBN.

3. EXPERIMENT

3.1 Experiment Settings

The experiment data(including click-through data, queries and documents) we use in this paper is all extracted from a Chinese business search engine (http://www.sogou.com). The labels, depicting the relevance between a query-document pair in the training set, are 5 levels from 0 to 4, and 0 means irrelevant and 4 means perfectly relevant, which are judged by professional annotators.

The click-through data is chosen from March 23^{rd} to March 30^{th} , 2924 queries and 1000 documents for each query are sampled and 87 ranking features are extracted for each query-document pair, such as $PageRank$ and $BM25$.

We use 4 strategies to generate labels in the training set when traning a ranking function with Learning to Rank method $Ranking\ SVM$ [5, 4], and compare the performance of ranking function based on these 4 strategies: HIL means that labels are judged by humans; UCM_ALL means that labels are predicted by the Unified Click Model with all 4 features in section 2.3; UCM_COEC means with three features of nCS, nRS and $COEC$; UCM_DBN means with three features of nCS, nRS, and DBN. Three measures of $P@N$, $NDCG@N$ and MAP are chosen to evaluate the performance of ranking functions learned through $Ranking\ SVM$ based on the training set, which is labeled by 4 strategies.

3.2 Experimental Results and Analysis

The comparison performance of the ranking functions learned through $Ranking\ SVM$ based on HIL, UCM_ALL, UCM_COEC, and UCM_DBN with the evaluation measures of $P@1, 5, 10$, $NDCG@1, 5, 10$ and MAP shows in Fig. 1. The experimental results show that UCM_ALL is the best method. We conducted t-tests on the improvements in terms of $NDCG@1 \sim 10$, and the results show that the improvements of UCM_ALL over HIL, UCM_COEC and UCM_DBN are statistically significant ($p - value < 0.01$). In three label-predicted methods, UCM_ALL is the only one that outperforms the HIL, and UCM_COEC performs better than UCM_DBN.

From the experimental results, the Unified Click Model performs best in predicting the label between a query-document pair, even outperforms the labels judged by humans. This might be the reason that the quantity of labels predicted by Unified Click Model is larger than ones judged by humans. Because the click-through data is very easy to obtain from search engines and contains the information that whether the users like the ranking for the URLs and which results the users prefer. So the Unified Click Model takes advantage of crowd wisdom in the click-through data.

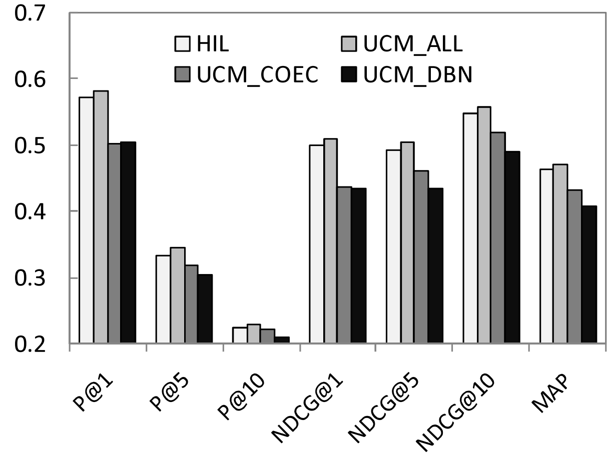

Figure 1: Performance Comparison with 4 Strategies Generating Labels

4. CONCLUSIONS

In this paper, we propose a Unified Click Model to predict the labels depicting the relevance between a query-documents pair. This click model could unify the advantages of Position Model and Cascade Model and the experimental results show that labels predicted by our model could replace the ones judged by humans and outperform the labels predicted by other click models.

5. ACKNOWLEDGMENTS

This work is supported by Natural Science Foundation (60736044, 60903107) and Research Fund for the Doctoral Program of Higher Education of China (20090002120005).

6. REFERENCES

[1] A. Broder. A taxonomy of web search. In *SIGIR Forum*, 2002.
[2] O. Chapelle and Y. Zhang. A dynamic bayesian network click model for web search ranking. In *WWW*, pages 1–10, 2009.
[3] N. Craswell, O. Zoeter, M. Taylor, and B. Ramsey. An experimental comparison of click position-bias models. In *WSDM*, pages 87–94, 2008.
[4] R. Herbrich and et.al. Large margin rank boundaries for ordinal regression. In *Advances in Large Margin Classifiers*, pages 115–132, 2000.
[5] T. Joachims. Optimizing search engines using clickthrough data. In *KDD 2002*, pages 133–142.
[6] T. Qin and et.al. Letor: A benchmark collection for research on learning to rank for information retrieval. In *Information Retrieval Journal*, 2010.
[7] D. E. Rose and D. Levinson. Understanding user goals in web search. In *WWW*, 2004.
[8] L. R. Yiqun Liu, Min Zhang and S. Ma. Automatic query type identification based on click through information. In *AIRS*, 2006.
[9] V. Zhang and R. Jones. Comparing click logs and editorial labels for training query rewriting. In *WWW 2007 Workshop: Query Log Analysis: Social And Technological Challenges*.