Proceedings of the ACM SIGIR Workshop on

# Information Retrieval and Advertising (IRA-2009)

held in conjunction with

**The 32nd Annual ACM SIGIR conference (SIGIR-2009)**

June 23, 2009

Boston, MA, USA


Workshop Organizers:

Mikhail Bilenko (*Microsoft Research*)

Evgeniy Gabrilovich (*Yahoo! Research*)

Matthew Richardson (*Microsoft Research*)

Yi Zhang (*University of California at Santa Cruz*)

## Organizing Committee

| | |
|---|---|
| Mikhail Bilenko | *Microsoft Research* |
| Evgeniy Gabrilovich | *Yahoo! Research* |
| Matthew Richardson | *Microsoft Research* |
| Yi Zhang | *University of California at Santa Cruz* |

## Invited Speakers

| | |
|---|---|
| Susan Athey | *Harvard University* |
| Yoelle Maarek | *Yahoo! Labs* |
| Kamal Nigam | *Google* |

## Program Committee

| | |
|---|---|
| Eugene Agichtein | *Emory University* |
| Susan Athey | *Harvard University* |
| Sugato Basu | *Google Research* |
| Andrei Broder | *Yahoo! Research* |
| Yiling Chen | *Harvard University* |
| Tao Chen | *University of Maryland, College Park* |
| Max Chickering | *Microsoft Research* |
| Silviu-Petru Cucerzan | *Microsoft Research* |
| Ewa Dominowska | *Microsoft* |
| Susan Dumais | *Microsoft Research* |
| Ayman Farahat | *Addkick* |
| Jon Feldman | *Google* |
| Rayid Ghani | *Accenture Technology Labs* |
| Anindya Ghose | *New York University* |
| Monika Henzinger | *Google* |
| Tao Hong | *Baidu.com* |
| Kartik Hosanagar | *University of Pennsylvania* |
| Nicole Immorlica | *Northwestern University* |
| Rong Jin | *Michigan State University* |
| Vanja Josifovski | *Yahoo! Research* |
| Oren Kurland | *Technion* |
| Ping Li | *Cornell Unversity* |
| Donald Metzler | *Yahoo! Research* |
| Mike Moran | *Freelance Consultant* |
| Vanessa Murdock | *Yahoo! Research Barcelona* |
| David Pennock | *Yahoo! Research* |
| Yan Qu | *Advertising.com* |
| Filip Radlinski | *Microsoft Research* |
| Hema Raghavan | *Yahoo! Labs* |
| Robert Ragno | *Specific Media* |
| James Shanahan | *Church and Duncan Group* |
| Dou Shen | *adLabs* |
| Arun Surendran | *Microsoft AdLabs* |
| Marcin Sydow | *Polish-Japanese Institute of Information Technology* |

| | |
|---|---|
| Ankur Teredesai | *University of Washington* |
| Roumen Vragov | *CUNY* |
| Michael Wellman | *University of Michigan* |
| Ryen White | *Microsoft Research* |
| Qiang Yang | *Hong Kong University of Science and Technology* |
| Sha Yang | *New York University* |

## Web Site

http://ira09.soe.ucsc.edu/

# Preface

The SIGIR Workshop on Information Retrieval and Advertising (IRA-2009) will be held in Boston, MA, USA on July 23, 2009. This second edition of the workshop is intended to bring together researchers and practitioners from academia and industry to discuss the latest developments in various aspects of online advertising and its interactions with information retrieval technologies. Our hope is that the workshop will serve as a lively, impactful forum for discussing new issues, as well as strengthening collaborations between industry and academia.

Online advertising has become the primary business model that supports a significant fraction of today's Web experience, including major Web search engines and numerous content-driven websites. Computational advertising systems employ many IR techniques alongside approaches developed in statistical modeling and machine learning, large-scale data processing, optimization, microeconomics, and human-computer interaction. Despite its commercial significance, computational advertising is a relatively young research discipline, which calls for cross-pollination of ideas and approaches from the different areas, which we hope the workshop will help promote.

We have attempted to choose a workshop format that would encourage lively discussion and active participation. To this end, we have compiled a schedule that includes a broad range of contribution formats: paper presentations describing latest research in online advertising, invited talks from industry and academia leaders from very different yet very relevant research areas, and, finally, position statements that we hope will promote debates, idea sharing and fruitful conversations in the concluding discussion session.

The six papers that have been selected for oral presentation cover a wide array of areas, from statistical analysis of advertising inventory reflecting world events to machine learning techniques for advertising selection. The paper presentations are to be followed by two position statements that point out new online advertising challenges related to classical IR issues: query processing and user interaction. Our hope is that this diversity of perspectives combined with relevance to core Information Retrieval topics will be appreciated by the attendees, as undoubtedly will the invited talks from Susan Athey, Yoelle Maarek and Kamal Nigam, who between them cover a fascinating range of expertise in fields that are foundational to online advertising: economics, information retrieval, and machine learning.

In closing, we would like to thank the authors who submitted the papers, program committee members for all their work in reviewing the submissions, as well as SIGIR 2009 workshop chairs for their support.

Misha Bilenko, Evgeniy Gabrilovich, Matt Richardson, Yi Zhang
*Organizing Committee*
July 2009

# Workshop Program

## Thursday, July 23, 2009.

| | | |
|---|---|---|
| 9:00 | **Welcome** | |
| 9:10 | **Invited Talk** | *Ads in Dynamic Query Suggestions* by Yoelle Maarek. |
| 9:50 | **Paper** | *Get more Clicks!* by Derek Hao Hu, Evan Wei Xiang and Qiang Yang. |
| 10:10 | **Paper** | *Sponsored Search for Political Campaigning during the 2008 US Elections* by Eni Mustafaraj and Panagiotis Takis Metaxas. |
| 10:30 | **Break** | |
| 11:00 | **Invited Talk** | *Online Advertising: Designing and Optimizing Marketplaces* by Susan Athey. |
| 12:00 | **Paper** | *The Effect of Some Sponsored Search Auction Rules on Social Welfare: Preliminary Results from an Exploratory Study in the Laboratory* by Roumen Vragov, David Porter and Vernon Smith. |
| 12:30 | **Lunch** | |
| 1:30 | **Invited Talk** | *High Precision Text Mining for Product Search* by Kamal Nigam |
| 2:30 | **Paper** | *Exploring Collaborative Filtering for Sponsored Search* by Sarah Tyler, Yi Zhang and Dou Shen. |
| 3:00 | **Break** | |
| 3:30 | **Paper** | *Towards Advertising in Social Networks* by Maryam Karimzadehgan, ChengXiang Zhai and Manish Agrawal. |
| 4:00 | **Paper** | *A Preliminary Study on Dynamic Keyword Extraction for Contextual Advertising* by Wen YE, Wenjie Li, Furu Wei and Chunbao Li. |
| 4:20 | **Position Talks** | *Better Query Modeling for Sponsored Search* by Hema Raghavan. |
| | | *Flexible Advertising Is Gaining Momentum* by Roumen Vragov. |
| 5:00 | **Informal discussion** | |
| 5:30 | **Workshop concludes** | |

# Table of Contents

## Invited Talks



## Position Statements



## Papers

# Ads in Dynamic Query Suggestions
# Extended Abstract

Yoelle Maarek
Yahoo! Labs Israel Ltd
MATAM, Advanced Technology Center
Haifa 31905, ISRAEL
yoelle_maarek@yahoo.com

Since the early days of Web search, the search "rectangle" kept its original simplistic form and most of the enhancements in terms of interaction with the user, concentrated on the results page. Search engines spent most of their efforts improving the search result page by adding a number of improvements such as spelling correction, snippets, sitelinks, translation link, etc. It is only recently that the attention has switched to the search box with the advent of dynamic query suggestion services such as Google Suggest or Yahoo! Search Assist. We started to see users interacting with the search box, selecting query suggestions being dynamically offered to them as they type. This paradigm involves new challenges as compared to regular search: First in terms of "effectiveness", these queries need to "look right" to users and relevant to information needs expressed by only a few characters. Second in terms of "efficiency", response time must be even faster that search response so as to follow the typing pace, and finally in terms of user experience, as this is the very front end, or the "first impression" the user gets from the search engine.

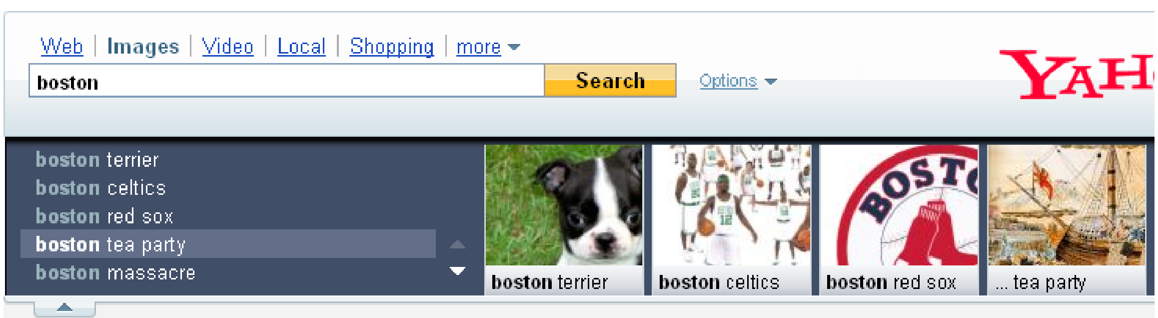


**Figure 1: Thumbnails in Yahoo! Search Assist**

In spite of these challenges, we have seen these suggestion services increase in reach and quality, going beyond query suggestions, in order to jump directly to some abbreviated results such as thumbnails in Yahoo! Image search as shown in the Figure above. An additional change occurred very recently when Google announced in its blog[1], possibly one of the most disruptive changes in this space: sponsored links (e.g., Ads) in the suggestion box as shown in the Figure below. In this talk, we will investigate the technical challenges involved in expanding the scope of the search box to include Ads along the three criteria mentioned before: effectiveness, efficiency and user experience. We will also discuss the advantages and risks of serving Ads in such settings and present our views on the disruptive potential of this change.

[1] http://googleblog.blogspot.com/2009/05/faster-is-better-on-google-suggest.html

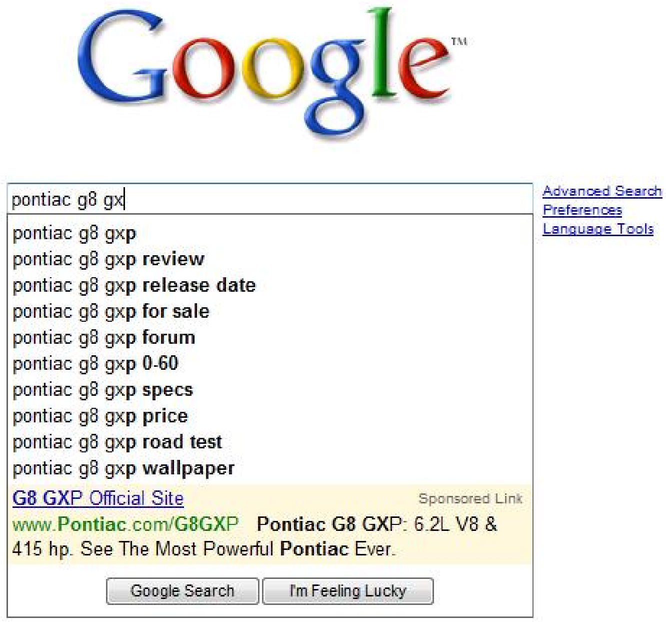


**Figure 2: Ads in Google Suggest**

**Bio:** Yoelle Maarek is a Senior Research Director at Yahoo! Labs in Haifa, Israel, which she joined in June 2009. Before this, she was Engineering Director at the Google Haifa Engineering Center, which she founded in March 2006 and grew to close to 40 engineers. Her team at Google Haifa launched one of the most visible features in Web search in the recent years: "Google Suggest", a query completion feature that is now deployed on google.com as well as a series of Google properties such as YouTube, iGoogle and Mobile Search, in more than 150 languages. Under Yoelle's supervision, the Google Haifa team launched several other features such as Searching Ads (See http://www.google.com/sponsoredlinks) and Interactive Annotations on YouTube (http://www.youtube.com/t/annotations_about). Prior to this, Yoelle had been with IBM Research since 1989. While at IBM Research, she held a series of technical and management appointments first at the T.J. Watson Research in New York, USA, and then at the IBM Haifa Research Lab in Israel until Feb 2006, where she contributed to IBM Enterprise search offerings. Her two last positions were Distinguished Engineer and Department Group Manager in the area of search and collaboration. She graduated from the "Ecole Nationale des Ponts et Chaussees" in Paris, France, and received her DEA (graduate degree) in Computer Science from Paris VI University, both in 1985. She was a visiting PhD student at Columbia University in NY in 1986/87. She received her PhD in Computer Science from the Technion, in Haifa, Israel, in 1989. Yoelle's research interests include information retrieval, Web applications, and collaborative technologies. She has published over 50 papers and articles in these fields. She served as chair or vice-chair of several technical tracks at the WWW conference series and as senior or regular PC member at most ACM SIGIR conferences in the last 10 years. She also chaired and moderated multiple workshops and panels at both WWW and SIGIR conferences. Most recently, she served as co-chair (with Andrei Broder) of the Panels track at WWW'2008 and as Technical Program co-chair (with Wolfgang Nejdl) at WWW'2009, that was held in Madrid in April 2009. Yoelle is also a member of the Board of Governors of the Technion, Israel Institute of Technology.

# Online Advertising: Designing and Optimizing Marketplaces

Susan Athey
Harvard University

The advent of online advertising has brought with it the need for new designs for markets that succeed in attracting participants and becoming viable businesses. Real-world design has been guided by theoretical insights, practical experience, and continual feedback from experimentation and data analytics. Statistical models have dual roles: they are an important part of the technology that ranks and delivers ads, and they are also used to evaluate the performance of the marketplace. This talk will focus on the interaction between theory and statistical analysis in paid search. It will highlight theoretical models of advertiser bidding behavior and market design that incorporate consumer search, as well as empirical models that help evaluate advertiser incentives and behavior. There can be important feedback between the design and performance of algorithms for scoring ads, and the incentives faced by bidders in online auctions as well as the efficiency of these auctions.

**Biography**

Susan Athey is a Professor of Economics at Harvard University. She received her Bachelor of Science degree from Duke University and her Ph.D. in Economics from Stanford University's Graduate School of Business. After teaching at MIT for six years and Stanford for five years, she moved to Harvard in 2006. Her current research focuses on auction theory, the design of auction-based markets, and the statistical analysis of auction data. She is an expert in a broad range of economic fields – including industrial organization, econometrics, and microeconomic theory – and has used game theory to examine firm strategy when firms have private information. She advises governments and businesses on the design of auction-based marketplaces, and she currently serves as a consultant for Microsoft Corporation in the role of Chief Economist.

In 2007, Professor Athey was named the first female recipient of the American Economic Association's prestigious John Bates Clark Medal, awarded every other year to the most accomplished American economist under the age of 40. She is a fellow of the American Academy of Arts and Sciences and the Econometric Society, and she serves as an elected member of the Council of the Econometric Society and the Executive Committee of the American Economics Association.

# High Precision Text Mining for Product Search

Kamal Nigam
Google

Product search is quite similar to online advertising in that the target audience and content providers have significant commercial intent. However, to provide a rich shopping experience product search requires understanding products not just as text but as structured data. These user interfaces are unforgiving of underlying data errors. The text mining and machine learning techniques used in product search must thus have unusually high precision. This talk will provide an overview of these different challenges and present details on two such applications.

**Biography**

Kamal Nigam is an Engineering Manager at Google Pittsburgh leading projects in product search, information extraction and data mining. He is also Adjunct Faculty in the Machine Learning Department at Carnegie Mellon University. His research interests lie at the intersection of text analysis, efficient use of human effort, and efficient use of unlabeled data. He received his Ph.D. from Carnegie Mellon University in Computer Science, and his S.B. from Massachusetts Institute of Technology. Prior to joining Google in 2006, he was Director of Applied Research at Intelliseek, a company applying text mining on web data for market research, and previously a Research Scientist at Whizbang Labs, a company specializing in information extraction on the web.

# Better Query Modeling for Sponsored Search

Hema Raghavan
Yahoo! Inc
4401 Great America Parkway
Santa Clara, CA, 95054
raghavan@yahoo-inc.com

## 1. INTRODUCTION

A primary difference between web search and sponsored search (SS) is that the web corpus is significantly larger than the ads database with more diverse content making it more likely to find relevant documents on the web using keyword-match techniques. Additionally users' are more tolerant to bad search results than to bad ads [3] and are more willing to reformulate queries when no relevant search results are found. In SS, however, there are many queries for which there are no relevant ads and standard IR techniques can result in spurious matches to unimportant words in the query. In such cases it is much better not to show any ads. These issues get particularly exacerbated for tail queries where *advance match* plays a dominant role. Yet, when possible, we do want to find relevant ads since a significant proportion of the SS revenue lies in monetizing the tail well. In this context we think that good query analysis can help achieve better accuracy for sponsored search in the tail. We describe 2 key problems in this regard. The ideas are however not restricted to SS but apply to web search as well.

## 2. KEY PROBLEM AREAS

### 2.1 Query Segmentation & Weighting

The first problem area is of "chunking" a query to determine key concepts in it, and then determining appropriate relative weights of these concepts. Consider the query "Donna Karan New York in Boston". The ideal segmentation is "Donna Karan New York | in | Boston". At the time of writing this paper all 3 major search engines (Microsoft, Yahoo! and Google) don't show the most relevant search result for this query at position one either for search or for SS. Google shows ads for Donna Karan products but the top search result is for "Barneys New York". The Microsoft engine also shows flights from New York to Boston in a sponsored listing and Yahoo! shows no ads. All engines show the wikipedia entry ranked above dkny.com. Additionally, for the query "donna karan new york" all 3 show the map of the store location in Manhattan on the search results page. (Note: The query "dkny" – the more popular form of the query – retrieves appropriate search results and ads on all engines). Clearly this is a problem query with spurious matches for "New York". Determining the correct segmentation and finding a good relative ordering of the segments' importance is key to solving this problem. A possible weight of importance could be modeled by the probability that the segment must appear in a relevant document $P(I_{seg})$. In this case if we could automatically determine via some machine learning method that $P(I_{seg=\text{Donna Karen New York}}) = 1$, we could aim to design a ranking function that enforces that "Donna Karen New York" appear in all retrieved documents (more in the next section). Some recent works [1, 4] address this problem by using machine learning algorithms to learn term weights for ranking. However, the area is new and has not been studied much for SS.

### 2.2 Ranking Ads

One way to use $I_i$ discussed above is in a ranking function of the following form which is was proposed by [2]: $P(Q|D) = \prod_{i \in Q} P(I_i)P(seg_i|D) + (1 - P(I_i))P(seg_i|C)$, where the index $i$ goes over detected segments ($seg_i$) in the query ($Q$) and $I_i$ is as above. $P(seg_i|D)$ can be modeled with the maximum likelihood probability of observing $seg_i$ in the document $D$ and $P(seg_i|C)$ is a smoothing probability. This model has couple of nice properties in that if $P(I_{seg}) = 1$ then documents without the presence of the segment will get a score of 0. In addition it allows for coordination level ranking: i.e., if all the segments have a score of $P(I_{seg}) \to 1$, then documents containing all the segments will rank higher than those containing only one (a property hard to guarantee in most TF-IDF like functions). Other functions with such properties and perhaps overcoming the limitations of the above one can be designed e.g., if one can learn $P(I_{seg=\text{dkny}}) = 1$ for our example query then one can implicitly incorporate query expansions.

## 3. CONCLUSIONS

In my presentation I will talk about why good query modeling and analysis is required for sponsored search, focusing on, but not restricted to the 2 areas above.

## 4. REFERENCES

[1] M. Bendersky and W. B. Croft. Discovering key concepts in verbose queries. In *SIGIR '08*.

[2] D. Hiemstra. Term-specific smoothing for the language modeling approach to information retrieval. In *SIGIR '02*.

[3] B. Jansen and M. Resnick. Examining searcher perceptions of and interactions with sponsored results. In *Workshop on Sponsored Search Auctions*, 2005.

[4] M. Lease, J. Allan, and W. B. Croft. Regression Rank: Learning to Meet the Opportunity of Descriptive Queries. In *ECIR-2009*.


*SIGIR 2009* Proceedings of the 32nd Annual ACM SIGIR Conference on Research and development in information retrieval

# Flexible Advertising Is Gaining Momentum

Roumen Vragov
Zicklin School of Business
Baruch College, City University of New York
17 Lexington Avenue, Box B 11-220
New York, NY 10010
roumen_vragov@baruch.cuny.edu

Advertising on the web is becoming increasingly interactive. Advertising companies are still debating whether this tendency would force them to give consumers control over the commercials they view. Traditionally advertising companies used customer surveys, customer purchase histories, customer demographics, and sophisticated data-mining techniques to determine which ads to show and what medium to use. Until TiVo came along, customers were never given a chance to choose directly which commercials they would like to watch themselves. Now the Internet has the technological capacity to provide each customer with his or her own customized ad stream. I call this new advertising approach *flexible advertising*. In my opinion the web advertising industry has a lot to gain if it adopts this new approach. Flexible advertising is better than traditional advertising because:

1. Flexible advertising improves customer satisfaction since the customer is in control of the advertising process and can pick commercials that the customer himself/herself considers relevant.

2. Flexible advertising alleviates privacy issues since many customers are uncomfortable revealing private information in surveys or allowing companies to save and later use their personal information to target them in the future with ad messages that are not necessarily relevant anymore.

3. Flexible advertising improves advertiser conversion rates and thus product manufacturer profits since ad relevance will improve and ad costs will decrease.

4. Flexible advertising provides more incentives to advertisers and product manufacturers to be more precise in the contents of their ad messages, since customers would be able to bypass those advertisements which they deemed misleading.

Overall, flexible advertising is beneficial to every party in the advertising industry: product manufacturers, advertisers, customers and therefore it should be adopted. It is only a matter of figuring out how to divide the gains from it among all these parties.

# Get more Clicks!

Derek Hao Hu
derekhh@cse.ust.hk

Evan Wei Xiang
wxiang@cse.ust.hk

Qiang Yang
qyang@cse.ust.hk

Department of Computer Science and Engineering
Hong Kong University of Science and Technology

## ABSTRACT

Sponsored search has become increasingly important due to the rapid development of Web search engines and pay per click (PPC) is amongst one of the most important advertising models search engines currently use. One of the key questions in sponsored search is that: Given a query or a substituted keyword, which ads should search engines display to the users in order to maximize their revenue? In other words, given a keyword, how can we choose ads out of a candidate list that will have higher click-through rates (CTR)? Previous works have attempted to estimate the CTR of ads via a query-independent perspective. In this paper, instead of predicting the CTR of ads, we will propose a new ranking-based approach to select ads that would have higher click-through rates via a query-dependent perspective. We first analyze some manually constructed heuristic rules that could be used to distinguish good ads from bad ones and then show how we could combine these rules into our ranking-based approach to reach our aim. Experiments on real-world datasets have confirmed the effectiveness of our proposed approach.

## Categories and Subject Descriptors

I.2.6 [**Artificial Intelligence**]: Learning; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

## General Terms

Algorithms, Experimentation, Economics

## Keywords

click-through rate, sponsored search, paid search, pay per click, ranking

## 1. INTRODUCTION

Internet-based advertising (also known as paid search, sponsored search etc.) has undoubtedly become one of the most popular ways of textual advertising these days. A large portion of search engine companies' revenue come from paid search. The market for Internet-based advertising has risen to $10 billion and will approximately reach $24 billion by 2013[1].

There are many Internet advertising models and in this paper, we focus our attention on the commonly used cost per click (CPC) model (sometimes also known as pay per click (PPC) model). In CPC model, advertisers only pay when a user actually clicks on an advertisement and then visits the advertisers' website. Such a CPC model is currently widely being used by major search engines like Google, Yahoo and Microsoft Live Search. Other major advertising models include cost per impression (CPM), where advertisers pay money according to how many times there ads are shown, and cost per action (CPA), where advertisers only pay when the users actually complete a transaction.

Click-through rate (CTR) is a way of measuring the success of an online advertising campaign. A CTR is obtained by dividing the number of users who clicked on an ad on a web page by the number of times the ad was delivered (impressions). For example, if an ad was delivered 100 times (impressions delivered) and one person clicked on it (clicks recorded), then the resulting CTR would be 1 percent. In CPC model, it is obviously seen that the objective for search engines is to "persuade" people to take actions, in other words, click the shown advertisements, based solely on the advertisement descriptions - usually a few, at most around 100, well-chosen words.

The main process of CPC-based online advertising is as follows. Users submit a query and search engines substitute the query to a given keyword that may match some advertisers' bid keywords. Then, the search engines would choose some ads (usually 1 to 5) to be displayed according to some ad ranking criteria, amongst which one of the most important criteria is to rank by expected revenue, i.e. the product of the advertisements' bid amounts and the advertisements' estimated click-through rates. Since advertisements' bid amounts are known in advance, such a ranking procedure could be reduced to estimate the advertisements' click-through rates. Previous research work [5] have also attempted to estimate the click-through rate for new ads via a machine learning-based approach, or more precisely, a regression-based approach, where the click-through rates learned are irrelevant of the issued queries.

However, it is noteworthy to see that, when estimating the click-through rate for ads, it is more reasonable to take

[1]http://www.emarketer.com/Article.aspx?id=1006644

into the user issued queries into account. Since it would largely affect whether users' would actually click the shown ads. Thus, in this paper, we propose a new approach to estimate the *query-dependent* click-through rates for advertisements, instead of the *query-independent* approaches as shown in previous research works [5]. Furthermore, instead of the traditional regression-based approaches to estimate the click-through rates directly for advertisements, in this paper we propose a ranking-based approach to estimate the ranking of advertisements and then directly choose the top-ranked advertisements. Such a motivation is influenced by the EigenRank approach [2] proposed to solve the collaborative filtering problem and we would discuss our motivation for adopting the ranking-based approach later in detail.

The remainder of this paper is organized as follows. In Section 2, we would briefly discuss some previous research works related to sponsored search, in particular we would discuss some previous research works in predicting CTR, ad ranking and generating / substituting advertisement keywords. In Section 3, we would briefly review some important terminologies and background knowledge in CPC model and CTR prediction. In Section 4, we would discuss some manually constructed tips to distinguish good ads from ads and we would perform some statistical analysis of the usefulness and the coverage of such tips on a real-world commercial search engine data set. In Section 5, we would discuss our ranking approach in detail and in Section 6, we would show our algorithm performance on the real-world dataset and validate the effectiveness of our approach. Finally in Section 7, we would conclude this paper and discuss some possible directions in which we could carry on our future research work.

## 2. RELATED WORKS

There were many important research works related to sponsored search. Richardson et al. [5] tried to solve the problem of predicting the click-through rates for new ads or rarely clicked ads. As described in the previous section, their approach aims to estimate CTRs *explicitly* and is independent of issued queries. Craswell et al. evaluated different models of user search result browsing to analyze the positional bias of click-through rates and found that cascade model predicts the observed bias accurately.

Ad ranking has also been studied from a variety of perspectives. Some theoretical computer scientists try to model the ad ranking problem as a online bipartite matching problem and used revealing LP to derive an optimal algorithm with a competition ratio of $1 - 1/e$ of this problem [3]. Pandey and Olston [4] modeled the advertisement ranking problem as a multi-armed bandit problem and studied the tradeoff between exploration and exploitation.

Another interesting area worth studying is the query substitution problem for sponsored search, that is, user issued queries are replaced by other generated queries to match the keywords bid by advertisers. Jones et al. [1] built a model for selecting between candidate substitution queries by using a number of features relating the query-candidate pair. Later, in [6], an active learning algorithm is used to select the most informative samples to train the query rewriting relevance model. There are many important research works on the rather broad area of sponsored search. However, it is beyond our scope and page limit to describe them in detail.

## 3. BACKGROUND KNOWLEDGE

In this paper, our objective is to let search engines automatically choose the "good" ads to display, given a specific user query. So what is a "good" advertisement? In previous sections, we had already given the definition of click-through rates. It is natural and a conventional manner to use CTR as a means of "scoring criteria" for each ad. However, in real-world situations, an advertisement is usually binded to several keywords, and even the same advertisement might have different CTR when the keywords are different. Therefore, it is not reasonable to forget the impact of keyword on CTR.

Thus, our problem can be formally defined as follows: Given a keyword $k_n$, we have a set of candidate ads $A_n = \{ad_{ni}\}$, each of which is a potential match to the keyword $k_n$. Thus, we want to select the top-ranked advertisements such that their click-through rates (given the keyword is $k_n$) are maximum in the set $A_n$.

## 4. TIPS TO WRITE GOOD ADS

With the development of World Wide Web and search engines, web sites that provide news and information about search engines and search engine marketing, such as Search Engine Watch[2], are emerging everyday. Articles on such web sites are usually published by domain experts or speculators, who could offer a good analysis for search engines and service development. In this paper, We first gather some expert tips from such authoritative sources, and then investigate how to use such tips to boost our CTR performance by rearranging the ads for their corresponding keywords.

### 4.1 Tips from Human Experts

**Ad Group Keywords:** It is required that the ad group's keywords should appear at least once in the ad, and it is much better if it can appear in the headline such that we could draw the user's attention.

**Speak Directly to Your Audience:** Ad readers would tend to feel better when they believe the ads are specially written for them. Therefore, using words like "you" and "yours" would make the readers feel you are directly offering service to them.

**Call Them to Action:** It is much better to use imperative verbs in your ad descriptions like "Get", "Shop", "See", "Find", "Buy" rather than phrases like "visit our site" or "click to see". It's best to call your readers to action and tell them what you want them to do.

**Create a Sense of Urgency:** The ad descriptions should try to create a sense of emergency in texts, e.g. let the readers believe they would suffer or fail to benefit if they don't act right away as told in the ads.

**Free is Good:** Trying to use free offers and explicitly mention them in your ad description would boost clicks and conversions a lot. Using free offers is quite beneficial, especially when your product or service is high-priced or rather complex, or the sales cycle is long.

**Flaunt:** If your product or service has many competitors, it would be better to underscore your advantages, e.g. using claims like "top ratings", "best-quality products", "maximize profits", etc. They could make your ads more attractive.

**Capitalize Every Word:** Often an ad with the first letter of each world capitalized often has a higher CTR than

[2]http://searchenginewatch.com/

a version of the same ad with lower-case letters. Capitalize individual words in the display URL also often boosts CTR.

### 4.2 Statistics from Historical Logs

We first categorize these expert tips into seven groups. For each tip, we filer the keyword-ad pairs $S_t$ which satisfy the $\text{Tip}_t$:

$$S_t = \{\langle k_n, ad_{ni}\rangle | ad_{ni} \in \text{Tip}_t\}$$

We define the coverage for $\text{Tip}_t$ as

$$\text{Coverage}(\text{Tip}_t) = \frac{|\{\langle k_n, ad_{ni}\rangle | \langle k_n, ad_{ni}\rangle \in S_t\}|}{|\{\langle k_n, ad_{ni}\rangle\}|}$$

Then we calculate the tip coverage and the average CTR over the satisfied ads for these keyword $k_n$:

$$\text{CTR}_{\text{Satisfied}}(k_n) = \text{average}(\{\text{CTR}(ad_{ni}) | \langle k_n, ad_{ni}\rangle \in S_t\})$$

We also calculate the overall average CTR for all the ads belonging to the keyword $k_n$:

$$\text{CTR}_{\text{Overall}}(k_n) = \text{average}(\{\text{CTR}(ad_{ni}) | \langle k_n, *\rangle\})$$

The effectiveness and coverage of some human expert tips are shown in Table 1. We can find that the CTRs of the ads which satisfy the tips are better than the average overall CTRs of the ads which belong to the same set of keywords. Some tips are able to boost the CTRs of the ads a lot, for example, Tip Set1, Set4, Set5 and Set6. However, we can also observe that the coverage of some powerful tips are quite low. For example, the first tip of Set4 and Set6 can only cover 7.5% of all the advertisements.

We try to use some thesaurus-based methods to improve the coverage of these tips. For example, we generalize the first and third tips of Set3 with an action verb list, and we also try to extend the first tip of Set6 with either a descriptive or a motivating adjective word list. We can find that, while the coverage of the extended tip is increasing, the improvement of the CTR drops. In the next section, we propose to use a ranking model to ensemble these effective but low covered tips together.

## 5. OUR RANKING MODEL

Previous works have attempted to estimate the CTR of ads via a query-independent approach. They propose to use a uniform regression model to predict the CTR of each ad independently. However, such a CTR prediction task is not essentially the objective of ad ranking for search engines. Search engines only need to pick out several top ranked advertisements for specific query keywords. Therefore, essentially we should focus on building a ranking model instead of a regression model. Such a motivation is influenced by [2], where the authors proposed a ranking model to predict the top ranked items for the collaborative filtering problem. Therefore, we plan to learn a ranking function to output the rank of the ads for each keyword instead of inferring their CTRs independently.

Assume that there exists an input space $X \in R^m$, where $m$ denotes the number of features. For our ad ranking problem, the features can be either bag of words (BOW) representation of the ads or the expert tips. The output space of ranks is represented by a set of ordered labels $r^* = \{r_1, r_2, ...r_q\}$, in which $r_1 \succ r_2 \succ ... \succ r_q$. Here $r_{ni} \succ r_{nj}$ denotes a preference relationship, and in our ad ranking task it means $ad_{ni}$ obtains better CTR than $ad_{nj}$ for keyword $k_n$. Suppose given a set of ranked ads belong to $N$ keywords $S = \{(\vec{x}_n, r_n^*)\}_{n=1}^N$, our objective is to learn a set of weights $\vec{w}$ via solving a constraint optimization problem:

$$\begin{aligned}
&\text{Minimize}: V(\vec{w}, \vec{\xi}) = \frac{1}{2}\vec{w}\cdot\vec{w} + C\sum \xi_{i,j,n} \\
&\text{Subject To}: \\
&\forall (ad_{1i}, ad_{1j}) \in r_1^* : \vec{w}\cdot\Phi(k_1, ad_{1i}) \geq \vec{w}\cdot\Phi(k_1, ad_{1j}) + 1 - \xi_{i,j,1} \\
&... \\
&\forall (ad_{Ni}, ad_{Nj}) \in r_N^* : \vec{w}\cdot\Phi(k_N, ad_{Ni}) \geq \vec{w}\cdot\Phi(k_N, ad_{Nj}) + 1 - \xi_{i,j,N} \\
&\forall i \forall j \forall n : \xi_{i,j,n} \geq 0
\end{aligned} \tag{1}$$

where $\Phi$ is a linear function of the feature vector $\vec{x}$:

$$\Phi(k_n, ad_{ni}) = \langle \vec{w}, \vec{x}_{ni}\rangle$$

Comparing with the regression model, which aims to optimize the prediction loss on ads' CTR individually, our ranking model minimizes the misordered ad pairs with respect to their keywords.

## 6. EXPERIMENTS

### 6.1 Data Sets

In this section, we evaluate the effectiveness of our ranking model with historical ad clickthrough logs collected by a commercial search engine. For the historical logs, we collect two-week log data from a commercial search engine over 1.7 million ads for 0.3 million keywords. The log records the query the user issued, the displayed ads id with position information, and the clicked ads id. We filtered out the ads of which the impression number is lower than 200. After that there left 35,000 ads for about 6,000 keywords. We sampled 90% keywords for training and used the remaining 10% for testing, and we carried out the experiments for 10 times. For both the ranking model and regression model, we used the implementation in the $SVM^{light}$ package[3].

### 6.2 Evaluation Metric

**Kendall tau distance** is a metric that counts the number of pairwise disagreements between two lists. The larger the distance, the more dissimilar the two lists are. We use the *normalized Kendall tau distance*[4] between the ground truth and the advertisement rankings predicted by our ranking model.

**Discounted Cumulative Gain(DCG)** is an increasingly popular metric for evaluating ranked results in information retrieval. Using a graded relevance scale of items in a search engine result set, DCG measures the usefulness, or gain, of a item based on its position in the result list. We use the *normalized Discounted Cumulative Gain*[5] as a measure of average performance of our ranking model for the ad ranking of different keywords.

### 6.3 Results

The first experiment demonstrates the effectiveness of the ensemble over the expert tips. We carried out three groups of experiments which consider 1) No Tip information, 2) Single

[3]http://svmlight.joachims.org/

[4]http://en.wikipedia.org/wiki/Kendall_tau_distance

[5]http://en.wikipedia.org/wiki/Discounted_cumulative_gain

**Table 1: Effectiveness and Coverage of Expert Tips**

| Tip Sets | Tip | Satisfied CTR | Overall CTR | Coverage |
|---|---|---|---|---|
| Set1 | Keyword in abstract | 0.083 | 0.067 | 36% |
| | Keyword in URL | 0.119 | 0.072 | 27.5% |
| Set2 | Contain {you,your} | 0.071 | 0.066 | 14.3% |
| Set3 | Contain {get,shop,see,find,buy,take} | 0.072 | 0.066 | 26.8% |
| | Contain words in action word list | 0.070 | 0.067 | 49.5% |
| | First word is {get,shop,see,find,buy,take} | 0.073 | 0.065 | 14.4% |
| | First word is in action word list | 0.068 | 0.066 | 23.1% |
| Set4 | Contain {now,before} | 0.078 | 0.065 | 7.4% |
| Set5 | Contain {free,zero} | 0.079 | 0.065 | 15.2% |
| | Contain price symbol $ | 0.075 | 0.062 | 3.6% |
| | Contain digital letters | 0.070 | 0.065 | 21.5% |
| Set6 | Contain {top,best,max,most,latest,newst} | 0.079 | 0.065 | 7.5% |
| | Contain words in descriptive word list | 0.078 | 0.066 | 18.9% |
| | Contain words in motivating word list | 0.078 | 0.067 | 27.9% |
| Set7 | Capitalize every word | 0.072 | 0.067 | 38.6% |
| | All words are short | 0.072 | 0.067 | 21.5% |

**Table 2: Ensemble of Expert Tips**

| Ensemble | Kendall Tau | NDCG | | | |
|---|---|---|---|---|---|
| | | Full | @1 | @3 | @5 |
| BOW | 0.637 | 0.930 | 0.725 | 0.867 | 0.897 |
| BOW+Single Tip | 0.640 | 0.931 | 0.727 | 0.869 | 0.898 |
| BOW+All Tips | 0.665 | 0.940 | 0.754 | 0.883 | 0.910 |

**Table 3: Different Ensemble Models**

| Ensemble | Kendall Tau | NDCG | | | |
|---|---|---|---|---|---|
| | | Full | @1 | @3 | @5 |
| Regression Model - SVR | | | | | |
| BOW | 0.630 | 0.928 | 0.703 | 0.862 | 0.892 |
| BOW+All Tips | 0.637 | 0.933 | 0.720 | 0.871 | 0.899 |
| Ranking Model - Ranking SVM | | | | | |
| BOW | 0.637 | 0.930 | 0.725 | 0.867 | 0.897 |
| BOW+All Tips | 0.665 | 0.940 | 0.754 | 0.883 | 0.910 |

Tip information, 3) All Tips information. The results are shown in Table 2. NDCG@1, NDCD@3 and NDCG@5 are also reported since we are more interested in the quality of the top ranked ads. We use the bag of words (BOW) of the ad content as the baseline feature representation which does not consider the expert tip information. "BOW + Single Tip" refers to the average result that adds only one expert tip to the BOW feature vector. "BOW + All Tips" ensembles all the expert tips with the original BOW features. We can find that "Single Tip" only boosts "BOW" a little, while our ranking based ensemble method outperforms the other two baselines sharply.

The second experiment shows that the ranking model can better capture the ads preference with respect to their corresponding keywords. We compare our ranking model with the regression model. We find that the performance of Support Vector Regression (SVR) is much worse than Ranking SVM. Moreover, if we incorporate all the expert tips using the regression model, the performance improvement is not significant over the BOW baseline. In contract, the ranking model can benefit more from the tips contributed by human experts.

# 7. CONCLUSION

In this paper, we investigate the problem of query-dependent ranking, i.e. given a keyword, how to choose ads out of a candidate list that will have higher clickthrough rates (CTR). Previous works have attempted to estimate the CTR of ads via a query-independent approach. In this paper, we will propose to select ads that would have higher clickthrough rates via a query-dependent ranking model. We first analyze some manually constructed heuristic rules that could be used to distinguish good ads from bad ones, and we find that the tips provided by human expert are quite effective by low value in coverage. Then we propose to combine these rules into our ranking-based ensemble model to reach our aim. Experiments on real-world data sets have confirmed the effectiveness of our proposed approach. In the future, we will investigate how to design a statistical model to generalize the expert tips to a wider range with the help of historical log data.

# 8. REFERENCES


[1] R. Jones, B. Rey, O. Madani, and W. Greiner. Generating query substitutions. In *WWW*, pages 387–396, 2006.

[2] N. N. Liu and Q. Yang. Eigenrank: a ranking-oriented approach to collaborative filtering. In *SIGIR*, pages 83–90, 2008.

[3] A. Mehta, A. Saberi, U. V. Vazirani, and V. V. Vazirani. Adwords and generalized online matching. *J. ACM*, 54(5), 2007.

[4] S. Pandey and C. Olston. Handling advertisements of unknown quality in search advertising. In *NIPS*, pages 1065–1072, 2006.

[5] M. Richardson, E. Dominowska, and R. Ragno. Predicting clicks: estimating the click-through rate for new ads. In *WWW*, pages 521–530, 2007.

[6] W. V. Zhang, X. He, B. Rey, and R. Jones. Query rewriting using active learning for sponsored search. In *SIGIR*, pages 853–854, 2007.

# Sponsored Search for Political Campaigning during the 2008 US Elections

Eni Mustafaraj
Department of Computer Science
Wellesley College
Wellesley, MA
emustafa@wellesley.edu

Panagiotis Takis Metaxas
Department of Computer Science
Wellesley College
Wellesley, MA
pmetaxas@wellesley.edu

## ABSTRACT

We have collected a set of 1131 textual ads that appeared in the Google Search results when searching for a candidate name running in the 2008 US Congressional elections. We have categorized the advertisers in four different categories: commercial, partisan, non-affiliated, and media. By analyzing the content of the collected ads, we discovered that the majority of them (63%) are commercial ads that have no political message, while the partisan group contributed only 14% of the ads. Furthermore, only 21 out of 124 monitored candidates were actively participating in sponsored search, by providing their own political message. We describe the different ways in which the advertisements were used and several problems that damage the quality of sponsored search, providing some suggestions to avoid such issues in the future.

## Keywords

sponsored search, content analysis, 2008 US elections

## 1. INTRODUCTION

The last report "The Internet and the 2008 Election" [1], published by Pew Internet & American Life Project, found out that 39% of Americans have used Internet to access "unfiltered" campaign materials during the 2008 primary elections. Since the search for information on the Web usually begins with queries in a search engine, the results produced by the search engine could have an impact on the kind of opinion an individual might form about a candidate. In the framework of a project aimed at capturing efforts of manipulating search engine results for political reasons, during a 6 months period (June – December 2008), we collected once a week search results targeting the names of more than a hundred candidates for the US 2008 Congressional elections, who were reportedly involved in crucial an highly contested races. Our analysis of the organic seach results has been reported in [2], thus, in this paper we will focus on the sponsored search results only.


*IRA '09* July 23, 2009. Boston, MA, USA


**Table 1: Distribution of ads during the 11-days period of data collection.**

| Date | Time (EDT) | # Ads | # Cand. |
|---|---|---|---|
| 2008-10-27 | 09:00 - 12:00 | 73 | 47 |
| 2008-10-28 | 15:00 - 18:00 | 65 | 48 |
| 2008-10-31 | 08:30 - 11:30 | 119 | 61 |
| 2008-11-04 | 11:00 - 14:00 | 119 | 71 |
| 2008-11-05 | 12:00 - 15:00 | 68 | 43 |
| 2008-11-06 | 08:15 - 11:15 | 73 | 50 |
| 2008-11-07 | 12:00 - 15:00 | 72 | 49 |
| 2008-11-11 | 12:00 - 15:00 | 80 | 57 |
| 2008-11-14 | 08:15 - 11:15 | 95 | 56 |
| 2008-11-24 | 11:15 - 14:15 | 176 | 84 |
| 2008-12-01 | 08:20 - 11:20 | 191 | 90 |

According to [3], the Google search engine attracts more than 70% of the query volume[1] in the United States, therefore, we have limited our data collection process to the results returned by Google. When using the Google API (Application Programming Interface) to automatically access the Google index in order to get results for a given query, the API will return the organic search results only. This is different from issuing a query to the Google web interface, because the latter will return an HTML page that usually contains several other types of results: advertisements at the top of the organic results, advertisements at the side of results, and interspersed in the page, results from searching news, videos, blogs, books, shopping, etc., or occasionally a list of related query search phrases in the bottom[2]. Since Google suggests the use of its API for automatic collection of search results, our weekly experiments contain only the organic search results. However, during the period Oct. 27 - Dec. 01 2008, in 11 different days, we automatically collected search results from the web interface, taking care to distribute the queries during a 3-hour period. The number of ads collected during these days, as well as the number of candidate names targeted by these ads are summarized in Table 1.

Reportedly [5], Barack Obama spent one million dollars on Google AdWords, in February 2008. Moreover, the indica-

[1]Although other sources report different numbers, all sources concur that Google is by a large margin the market leader in search.

[2]On May 12, 2009, Google announced another change to its Search service, called Search Options [4]. Therefore, the above description might not apply anymore to the appearance of search results in a web page.

tions for the gubernatorial races taking place in Fall 2009 are that the campaigns and issue groups are making an earlier and more sophisticated use of AdWords than the presidential candidates at the comparable stage [6]. Thus, AdWords is already becoming an important tool for online political campaigning, with the potential to surpass other types of advertisements in the future, because users are also moving en-masse toward online political information gathering. From this viewpoint, we think it is of value to look retrospectively on how sponsored search performed during the 2008 elections, in order to uncover potential issues that need to be carefully examined in the future. To our best knowledge, such an exploratory analysis has not appeared previously in the research literature. We recognize that our methodology for data collection has flaws, therefore, we suggest how to improve this process in the section of future work.

## 2. THE DATA COLLECTION PROCESS

During a careful selection process, described in [2], we had selected 60 congressional races that were predicted to be highly contested. Later on, for the purpose of collecting advertisement data, we added the two pairs of candidates in the presidential race. Thus, there were 124 candidates (120 candidates for the Congress and the 4 candidates for president/vice president) followed by our experiment. Each candidate name (in quotation marks) was issued as a query to the Google Search Web Interface, and the returned HTML pages containing organic and sponsored search results were stored. By parsing the HTML code, it was possible to detect and extract sponsored search results, when present. In total, we collected 1352 HTML files (during the first 3 days of collection, the 4 presidential candidates were not included, $3 * 120 + 8 * 124 = 1352$), and by parsing these files we extracted 1131 advertisements. By aggregating all extracted data, it resulted that 112 of the 124 candidates did have at least one advertisement targeting their name .

Differently from advertisements within websites, which are usually visually stimulating, sponsored search ads are purely in textual form. Commonly, a textual ad is a triplet containing a caption, a short message, and a URL. When using Google Search, textual ads are displayed either in the top of the page (right above the top ten list of results) or in the side. Google Search displays up to 3 ads in the top section and up to 8 ads in the side section. We counted 72 ads in the top position compared to 1059 ads in the side position. This fact shows that the quality score used by Google to determine an ad positin has ranked as non-top quality the majority of the ads. In the following section, we discuss several findings of our exploratory content analysis.

## 3. FINDINGS

By extracting the captions and URL portion of each textual ad, we found out that in the collected set of **1131** ads, there are **489** unique ads (i.e., their caption is unique) contributed by **147** websites (advertisers). After analyzing the content of the ads and the related URLs (we visited several websites, whose nature was not obvious from the ad text), we grouped the websites in four categories: **commercial** (websites that sell products or information), **partisan** (websites promoting or opposing candidates and their agendas), **non-affiliated** (websites contributing political and electoral information, but not affiliated with the candidates or their parties), and **media** (websites of newspapers, TV stations, magazines, etc.). Statistics about these four groups are summarized in Table 2. As the numbers indicate, 63% of the ads are purely commercial and have no political meaning. The following subsections discuss the content of ads for each category.

### 3.1 The Partisan Category

The analysis of the partisan category showed that only 21 candidates (17 running for the congress and the 4 for the presidency) had actively included ads directing to their campaign websites, with text messages such as:

> Support Senator John Sununu. Join Team Sununu Today!
> Darcy Burner is exactly what we need in Congress!
> Join John McCain's Team and Contribute Today! Thank You
> Tom Udall needs your help to keep fighting for New Mexico.

Besides these direct, positive ads, there are two other types of ads.

#### *3.1.1 Targeting the Opponent Name*

Some candidates had bought ads that targeted the name of their opponent. For example, when the name of Gerry Connolly was searched for, an ad from his opponent appeared, with the caption "Keith Fimian for Congress". One other example was that of the democratic candidate Tom Udall, whose ad "Tom Udall for New Mexico" appeared when searching for the republican candidate who was running for the house seat in his district, Darren White, although Tom Udall himself was racing for the senate seat. While these examples can be labeled as "awareness ads" (giving to people searching for a candidate the possibility to know about the opponent), there were occasions of negative ads targeting the opponents, although not directly from the candidates. For example, when searching for the candidate Norm Coleman, several negative ads directed to his opponent Al Franken would appear, such as "Franken: Unfit for office" or "Is Al Franken lying?", contributed by a website named `www.FranklyFranken.com`.

#### *3.1.2 Negative Ads*

The more frequent use of negative ads was to directly target the name of a candidate. We found 25 negative ads that have appeared 67 times, targeting 16 candidates. As negative ads we treated those ads that contained disapproving language toward the candidate. Some examples of negative ads are shown in Table 3. However, only 4 of these ads (for a total count of 11 appearances) were published by the candidates' websites, all the other negative ads came from other websites, not directly affiliated with the candidates. The most worrisome fact about the negative ads is that they originate from websites which were shut down directly after the elections, and that have used the candidates' names in their URLs. Some of those now extinct websites are:
`ShaheenForSenate.com`, `VoteMcNerneyOut.com`,
`GilibrandUnfiltered.com`, `TheRealBobRoggio.com`.
Their sudden disappearance might be an indicator of unfair and unethical political campaigning.

### 3.2 The Non-affiliated and Media Categories

Non-affiliated sites use the names of the candidates to attract traffic to their sites, by using template ads such as:

Table 2: **The four categories of websites, which ads were displayed in response to candidate names.**

| Website type | # sites | # targ. cand. | # ads | % |
|---|---|---|---|---|
| *Commercial* (e.g., `amazon.com`, `public-records-now.com`) | 79 | 92 | 712 | 63% |
| *Partisan* (e.g., `BarackObama.com`, `dccc.org`) | 37 | 39 | 157 | 14% |
| *Non-affiliated* (e.g., `theMiddleClass.com`, `houseRaceTracker.com`) | 16 | 72 | 206 | 18% |
| *Media* (e.g., `newser.com`, `FoxNex.com`) | 15 | 20 | 56 | 5% |
| **Total** | 147 | 112 | 1131 | 100% |

**Table 3: Examples of negative ads. Ad text, including errors, appears verbatim.**

| Ad Caption | Ad Text |
|---|---|
| Who's the real Bob Roggio | Toxic Chemicals. 500 PA Jobs Lost. Now Wants $2,031 in Higher Taxes! |
| McNerney Undermines War | A Sellout Democrat for Al Qaeda Apposed Reinforcements in Iraq |
| Women Against Sarah Palin | Anti-choice, anti-gay, pro-drilling Palin does not speak for women. |
| Lobbyist Steve Stivers | Career Lobbyist Steve Stivers for Congress |

View Rep. [Candidate Name] 's middle-class voting record here.

Do you agree with [Candidate Name] on the issues that matter to you?

For example, the website `themiddleclass.com` alone targeted 45 candidate names for a total of 110 ad occurrences, with the majority of them (around 75%) past the election date. The non-affiliated sites together targeted 72 candidate names for a total of 206 ad occurrences.

Media websites also use the same strategy, by using the candidate names in the caption of the ad, for example:

[Candidate Name] Fox-NEWS *provided by* `foxnews.com`

[Candidate Name] News *provided by* `examiner.com`, `news.aol.com`, etc.

The media category was the smallest one in the dataset. It had 15 different advertisers, targeting 20 candidate names for a total of 56 ad occurrences.

## 3.3 The Commercial Category

The commercial ads outweigh by far all other types of ads. In fact, we counted 712 occurrences of ads targeting 92 candidates. The first reason for the abundance of the commercial ads is that three commercial websites, specialized in finding people:

`public-records-now.com`
`usa-people-search.com`
`wink.com`

have contributed together 380 ad occurrences, targeting 71 candidate names. The second reason is that several candidates have names shared by other individuals that have monetizable professions. That happened to the candidate Brian Davis that had an artist doppelganger attracting 59 ads from online sellers of art, to the candidate Dan Seals that had a country singer doppelganger, or to Steve Greenberg that had a record producer doppelganger. Furthemore, some of the candidates themselves have in their course of life produced books or music, played professional sports, etc., so that websites such as `Amazon.com` and `eBay.com` will target their names. And yet another source of ads were sites trying to cash in with products not authorized by the candidates, as the examples in Table 4 show.

Very often, an ad will not target any of the candidates in a race. In fact, 25% of all ad occurrences were such mistargeted, commercial ads, because they used various spellings of the candidate names. The most common use of such ads were from the people finder websites, which will target any variation of a name and surmane, or even have ads targeting only the first name, such as: "Find Dean". While this kind of ads at least understands that the user is searching for a person name, there are other ads that miss the semantic category of the query. For example, when searching for the candidate Baron Hill, appeared the ad: "Barton Hills home 4 Sale", where Barton Hills is a geographic location; when searching for Darren White, appeared the ad: "White Pages directory", when searching for Victoria Wulsin appeared ads from the Victoria's Secret company, when searching for Nancy Boyda appeared the ad "Local Nannies", and when searching for Mark Begich appeared several ads related to companies or products with the keyword "mark", such as "Vanmark Collectibles", "TriMark Corporation", "Primemark", or "Colormark at Amazon.com". An interesting fact is that although there are five candidates with their first name Mark, only one of them, Mark Begich, attracted the mentioned ads.

# 4. DISCUSSION AND SUGGESTIONS

As in traditional media, the ads appearing in the sponsored search results were used to both promote or oppose a candidate's political message. However, there is a big difference. Advertisements in TV, radio, and newspapers clearly indicate who has paid for it, by mentioning this fact in the ad, so that voters know who is responsible for the message. Political online advertisements has yet to be regulated by law, however, the Federal Electoral Committee (FEC) in 2006 has advised [7] the following about Internet advertisements:

> Because Internet advertisements are public communications, an individual or group must include a disclaimer on any Internet advertisement that expressly advocates the election or defeat of a clearly identified Federal candidate, or on any Internet advertisement that solicits contributions.

None of the ads we collected and analyzed had any disclaimer in their text. Often several clicks in the landing page are needed to find out who is responsible for the advertised

Table 4: Examples of commercial ads.

| Ad Caption | Ad Text |
|---|---|
| Obama 4-Coin Collection | Tribute to 44th President of US in velvet gift box. Perfect gift! |
| President Obama T-Shirts | Yes We Did! Welcome Mr. President. Tees, Stickers, Buttons, Yard Signs |
| Barack Obama | Shop our Best Designs or Create Your Own Barack Obama Merchandise! |
| Palin Humor Tees | Get Sarah Palin Parody Shirts, Stickers, Buttons, & Gear |

message, and in the occasion of extinct websites, this is not possible at all. While it is true that for such short ads is difficult to have an extra disclaimer message, we believe that it might be possible to circumvent this drawback, for example, by including a small icon that indicates a pro or contra message.

Additionally, there are several other improvements that search engines can undertake in order to make sponsored search results more effective and help voters to avoid confusion or ambiguity. Given the fact that elections take place in two-year intervals and candidate names are known several months before the election date, there is plenty of time for search engines to put such measures into actions.

1. Maintain a list of candidate names registered with FEC. For searches matching names in this list, avoid displaying ads that do not contain the correct spelling of the candidate name either in the caption or in the text message. In this way, all spurious ads resulting from broad matching are eliminated.

2. Require a disclaimer for all ads targeting the names of political candidates.

3. Resolve cases of several individuals with the same name (similar to the Wikipedia disambiguation page for people's names).

4. Clearly mark political ads (for example by using a background color) to distinguish them from commercial or informational ads.

We think that such measures are in the interest of search engines. By taking actions to ensure that only relevant ads are delivered, more candidates, political groups, or others might be interested to advertise. By providing recognition cues and categorizing ads properly, the users might be encouraged to click more often on such ads.

# 5. FUTURE WORK

Our data collection methodology has some flaws. We searched for less than one fourth of the candidates (though that subgroup was involved in highly contended races); we collected data only for a few days before elections, only once a day, and for one geo location only (that of our server). Furthermore, our queries contained the names of the candidates only, whereas the candidates might have been advertising with many other words, such as "Illinois candidates for senate", "elections in Montana", "economic crisis", "war in Iraq", etc. Coming up with a comprehensive set of keywords in advance is something that requires much more preparation work, might need the involvement of a political analyst, and also a constant monitoring of the regional query trends volume using the Google AdWords tools.

Another important area of future study is that of contextual advertisements, which appear on third-party websites via, for example, the Google AdSense technology. In our corpus of 65,700 HTML pages, collected by storing the top 20 search results returned by Google when searching with the candidate's names over a 6 month period, 20% of the pages have in their HTML code the *google_ad_client* script that delivers the ads. Because the ads are delivered in real time and are not part of the HTML source of the page, we were not able to analyze their content. We need thus means to automatically record ads delivered within a webpage at a specific point in time, in order to analyze their content. Even for the candidates themselves it is important to know on what websites their advertisements are being displayed, because they might not agree to become affiliated with websites that use derogatory language. In particular, John McCain's campaign did withdraw several of his ads from some anti-Obama websites, when it was discovered that such websites were using offensive language toward Obama [8].

All these issues need to be addressed carefully, if a more comprehensive analysis of online advertisement for political campaigning is desired.

# 6. REFERENCES


[1] Online Report, *The Internet and the 2008 Election*, `http://www.pewinternet.org/Reports/2008/The-Internet-and-the-2008-Election.aspx`

[2] P. Metaxas and E. Mustafaraj, *The battle for the 2008 US Congressional Elections on the Web*. Proc. of WebSci-09: Society On-Line, March 18-20, 2009, Athens, Greece.

[3] StatCounter - Global Stats. Last access on July 10, 2009. `http://gs.statcounter.com/#search_engine-US-daily-20080701-20090610`

[4] The Official Google Blog, *More search options and other updates from our Searchology event*, May 12, 2009. `http://googleblog.blogspot.com/2009/05/more-search-options-and-other-updates.html`.

[5] S. L. Stirland, *Our Brand Is Crisis: Prez Candidates Buy Words To Brand Each Other Online*, Sep. 19, 2008. `http://www.wired.com/threatlevel/2008/09/our-brand-is-cr/`

[6] J. A. Vargas, *Major front in Virgina race is online*, Apr. 29, 2009. `http://www.washingtonpost.com/wp-dyn/content/article/2009/04/28/AR2009042803419.html`

[7] Online Document, *FEC Internet Rulemaking - Background and FAQ*, Mar. 27, 2006. `http://www.fec.gov/members/weintraub/nprm/statement20060327.pdf`

[8] P. Hamby, *McCain campaign pulls ads from some anti-Obama web sites*, Jul. 1, 2008. `http://www.cnn.com/2008/POLITICS/07/01/mccain.ads/`

# The Effect of Some Sponsored Search Auction Rules on Social Welfare: Preliminary Results from an Exploratory Study in the Laboratory

Roumen Vragov
IS Behavioral Research Laboratory, CUNY
1 Bernard Baruch Way
New York, NY 10010
646-239-6561
rumvra@yahoo.com

David Porter
Chapman University Endowed Chair in Experimental Economics
One University Dr.
Orange, CA 92866
714-997-6815
dporter@chapman.edu

Vernon Smith
Chapman University School of Law
One University Dr.
Orange, CA 92866
714-628-2830
vsmith@chapman.edu

## ABSTRACT

When consumers search sponsored links provided by a search engine they interact with advertisers in two distinct but related markets: the market for ads, and the market for the advertised products. The purely theoretical exploration of such complex combinatorial markets is limited because it requires assumptions about consumer and advertiser behavior that are too strict. This study explores the effects of some sponsored search auction rules on consumer surplus, advertiser profits, and search engine revenues through the use of laboratory experiments with human subjects. We find that, from the options we explored, the best payment method is pay-per-click and the best way to rank ads is by past click-through rates. We also suggest ways to extend the experimental design further to explore other important parameter spaces.

## Categories and Subject Descriptors

## General Terms

Economics, Experimentation

## Keywords

sponsored search, combinatorial auctions, experimental economics, web advertising

## 1. INTRODUCTION

Sponsored search on the Internet combines the characteristics of two complex problems in information economics. The first problem is that of optimal consumer search. The second problem is that of combinatorial auctions. In essence, sponsored search auctions are combinatorial in nature because the commercial value to an advertiser of a search phrase might be different than the sum of the individual values of the words in that phrase. In addition, the value of a phrase depends semantically on its word components, and the value of a position in a list of sponsored links depends on the number and position of other ads listed there. The value of an advertisement to a consumer is also combinatorial in nature because the value of one advertisement might depend upon the number and nature of the other advertisements currently displayed on the page. The purpose of this study is to create an environment in the laboratory that represents the essence of these two basic problems above, and then, using a standard 2x3 factorial design, answer the following research questions:

1. How do consumer surplus, seller profits, and search engine revenue differ when sponsored search results are ranked by click-through rate (CTR) versus by Relevance (R)?
2. How do consumer surplus, seller profits, and search engine revenue differ when sellers pay per click (PPC) versus when sellers pay per transaction (PPT)?

To accomplish this goal we adapt a version of a sponsored search auction as described in Vragov & Levine (2007), which already takes into consideration combinatorial values for consumers, to web search and also incorporate in it combinatorial values for advertisers. Since such a mechanism is too complex to model theoretically, we compare the effects of the auction rules mentioned above with the help of laboratory experiments with human subjects (see Smith, 2003). Through laboratory experiments we can take into account actual human behavior (which is often different from that prescribed by theory), and we can also observe actual mechanism practicality and ease of use. In addition, laboratory experiments provide a precise and easy way to measure consumer surplus and advertiser profits directly (see Smith, 1976).

## 2. LITERATURE REVIEW

So far there have been relatively few theoretical models of sponsored search auctions that take into account not only the way advertisers bid and the way their ads are ranked but also the prices that advertisers charge consumers once consumers find the products they need. Such models necessarily have to study complex relationships between two separate but related markets: the market for ads and the market for the advertised products. Such an endeavor usually requires quite strict assumptions about consumer search behavior and advertiser pricing and bidding strategies, but these attempts are necessary in order to provide answers to the research questions posed above.

The simplest model of the interaction between the two related markets is presented in Vragov (2009). According to his model there is no difference in consumer and advertiser surplus when advertisers pay per click versus when they pay per transaction. However, search engine revenue is higher in the latter case. The effect of ranking by relevance versus click-through is only considered in duopoly circumstances, and then, obviously, ranking by click-through might prevent a lower cost firm to enter the ad market. Athey & Ellison (2008) discuss a much more general and involved framework, which allows them to formulate two results that are related to our research questions. The first result is that click through weighting of bids does not cause differences in surplus in the limit. However, when the number of firms is small, click-through weighting causes inefficiencies in the set of listed firms and also inefficiencies in the ordering of listed firms. They also argue that differences in method of payment for advertisers are related to the informative content of the displayed ads. This implies that in absence of informative content there should not be differences in surplus or profits that depend on the method of payment. These results mimic the ones by Vragov (2009) reported earlier.

Blumrosen et al. (2008) conduct some interesting simulations using a theoretical model and observations from real advertising data. They find that pay per transaction is better than pay per click for the advertiser and the search engine. Next, Dellarocas and Viswanathan (2008) provide a theoretical model of the interaction between advertisers, search engine, and consumers and find that pay per transaction leads to higher prices and decrease in pay-offs for all. They also find that ranking advertisers by a product between bid and click through rate will improve all participants' pay-offs.

The last related model we mention here is the one by Thompson and Leyton (2008). The authors use innovative algorithms to compute that unweighted pay-per-impression auctions outperform weighted pay-per-impression auctions while weighted pay-per-click auctions outperform unweighted pay-per-click auctions. Also, when bids are discrete, VCG revenue is not a lower bound on the revenue of weighted pay-per-click GSPs. They also show that unweighted pay-per-impression auctions outperform weighted pay-per impression auctions while weighted pay-per-click auctions outperform unweighted pay-per-click auctions.

The literature review above tells us that results are often controversial which might indicate that results are not too robust to changes in assumptions, which are usually too strict. We decided to use laboratory experiments with human subjects to bring theoretical models mentioned above closer to reality and to investigate our research questions for the first time with one more method of scientific exploration. Comparing results from different methodologies can provide a fuller picture of the complex environments we are studying. In the experimental design described below we make no assumptions about consumers and advertisers (they are randomly picked from the undergraduate population of a large urban business school) and we use a very general product space that contains both substitutes and complements. The strictest assumptions we make is that the search engine has a perfect knowledge of ad relevance based on the search terms consumers type and that consumers know what to type in order to indicate how relevant an ad is to them. As discussed in the last section we plan to relax these assumptions as well in future designs.

# 3. EXPERIMENTAL ENVIRONMENT

Our experimental environment is symmetric because it consists of 6 buyers (consumers) and 6 sellers (advertisers). First we describe the parameters related to sellers. Every seller produces a unique product and every seller can sell up to 6 units of the product at constant marginal cost of production. The costs were chosen randomly from a uniform distribution with support [0, 100] and rounded up to the nearest 5. There are two product categories: A and B. There are three products in category A: A1, A2, A3, and three products in category B: B1, B2, and B3. Each product is uniquely assigned to be produced by only one seller. Table 1 shows the product assignment and the per-unit cost of production. Values and costs are shown in experimental dollars.

**Table 1. Supply features**

| Seller | Product | Cost per unit | Production Capacity |
|---|---|---|---|
| 1 | A1 | 55 | 6 |
| 2 | A2 | 50 | 6 |
| 3 | A3 | 40 | 6 |
| 4 | B1 | 35 | 6 |
| 5 | B2 | 25 | 6 |
| 6 | B3 | 45 | 6 |

We next describe the parameters on the demand side. As we mentioned before we have 6 buyers with combinatorial preferences. The values and costs for each buyer are shown in Table 2. The values for Buyer 1 were chosen randomly from a uniform distribution with support [50, 150] then rounded up to the nearest 5 and sorted from highest to lowest. The remaining values were chosen in such a way so as to introduce some variation in the demand environment in terms of the slopes of the demand curves, the rakings of products for buyers in different groups and search costs.

**Table 2. Buyers' values, bonus, and click costs.**

| | **Group 1** | | **Group 2** | | **Group 3** | |
|---|---|---|---|---|---|---|
| **Product** | **Buyer 1** | **Buyer 2** | **Buyer 3** | **Buyer 4** | **Buyer 5** | **Buyer 6** |
| A1 | 120 | 120 | 110 | 110 | 90 | 70 |
| A2 | 100 | 100 | 120 | 120 | 105 | 95 |
| A3 | 80 | 80 | 115 | 100 | 120 | 120 |
| B1 | 65 | 65 | 65 | 65 | 65 | 65 |
| B2 | 80 | 80 | 80 | 80 | 80 | 80 |
| B3 | 75 | 75 | 75 | 75 | 75 | 75 |
| Bonus (1 A & 1 B) | 30 | 30 | 30 | 30 | 30 | 30 |
| Click Cost | 5 | 15 | 5 | 5 | 5 | 15 |

Buyers are divided in three equal groups (2 subjects each) depending on which product from Category A they like best. Buyers 1-2 prefer A1, Buyers 3-4 prefer A2, and Buyers 5-6 prefer A3. In addition buyers are distinguished in terms of their click costs. Buyers 1, 3, 4, and 5 have a click cost of 5; Buyers 2 and 6 have a click cost of 15. For some buyers the differences in values for the products in category A are different. For example Buyer 1 and Buyer 2 share the same differences among product values. For Buyers 3, 4, 5, and 6 the differences between values for products in Category A are respectively 5, 10, 15, and 25. The values for the products in Category B are the same for all buyers. The bonus that buyers receive when they buy one product from Category A and one product from Category B is also the same for all subjects.

## 4. MARKET MECHANISM

When an experimental round starts first it is sellers' turn to make their decisions. Sellers have to decide how much to charge buyers who wish to buy their product and how much to pay to the search engine for being displayed on the buyers' screen. Sellers bid for exposure in our experimental design. The three sellers in Category B each have to submit a price for their product and a bid to the search engine. The three sellers in category A have more options. Since buyers differ in terms of their preferences for products in Category A, sellers are allowed to submit bids separately for buyers who prefer A1, buyers who prefer A2, and buyers who prefer A3. Thus sellers who produce products in Category A can submit a price for their product and up to three bids: one bid for product A1, A2, and A3. If they submit the same bid for all products in category A, this means that they are bidding on one more general search phrase that describes all products in category A. If the three bids they submit are different, then sellers chose to bid on three more specific search phrases that more specifically describe each of the products in the A category.

After sellers have made their decisions, the search engine collects all bids and prices and decides separately which products to display to buyers who prefer respectively A1, A2, and A3 and also in what order to display them. The search engine picks the sellers with the top 4 bids for each of the three product groups and displays their products to the buyers.

**Table 3. Example of sellers' bids and prices**

| **Seller** | **Bid for Buyers in Group** | **Bids** | **Prices** |
|---|---|---|---|
| 1 | 1 | 11 | 90 |
| | 2 | 5 | |
| | 3 | 1 | |
| 2 | 1 | 3 | 105 |
| | 2 | 7 | |
| | 3 | 9 | |
| 3 | 1 | 4 | 99 |
| | 2 | 7 | |
| | 3 | 8 | |
| 4 | Same for all | 10 | 60 |
| 5 | Same for all | 8 | 62 |
| 6 | Same for all | 3 | 53 |

***Example:*** *Given the sellers' bids displayed in Table 3 we will demonstrate how the search engine decides which products to show to the various groups of buyers. Let us first run the procedure for Group 1 (Buyer 1 and Buyer 2). Buyers 1 and 2 prefer A1. We take the bids from sellers 4, 5, and 6 and also the bids for A1 from sellers 1, 2, and 3 and align them from highest to lowest. The resulting array of bids is shown in Table 4, column 2. A similar procedure is performed for the remaining two groups of buyers (see columns 3 and 4).*

**Table 4. Ranking of bids from Table 3.**

| | **Buyers in Group 1** | **Buyers in Group 2** | **Buyers in Group 3** |
|---|---|---|---|
| Shown | A1 – 11 | B1 – 10 | B1 – 10 |
| Shown | B1 – 10 | B2 – 8 | A2 – 9 |
| Shown | B2 – 8 | A2 – 7 | A3 – 8 |
| Shown (Lowest Accepted Bid) | A3 – 4 | A3 – 7 | B2 – 8 |
| Not shown | B3 – 3 | A1 – 5 | B3 – 3 |
| Not shown | A2 – 3 | B3 – 3 | A1 – 1 |

The order in which the top 4 bids are shown to each buyer varies by treatment. When the ranking is done, buyers can proceed to make their decisions. Each buyer can see that four of the products are available for sale. A buyer can click on a product to check its price, which results in a click cost as shown in Table 2. After checking some or all of the prices, the buyer can proceed to make purchase decisions. If a buyer decides to purchase a product, the buyer receives the value of the product from Table 2 as revenue and has to pay the price that the seller of that product indicated.

***Example (con'd)****: Suppose that Buyer 1 checks the prices of three of the four products displayed to her: A1, B1, A3 (see Table 4, column 2). The prices of these products are respectively: 90, 60, 99 as submitted by the sellers (see Table 3, column 5). Given these prices, Buyer 1 decides to buy A1 for 90 and B1 for 60. Using the values and costs in Table 2, we can calculate Buyer 1's profit in this round. It is equal to 120 + 65 + 30 – 90 – 60 – 3 x 5 = 50. Note that Buyer 1 receives a bonus of 30 because she bought a product from category A and a product from category B. Buyer 1 also incurs a click cost of 15 because she clicked three times to check the prices of three products.*

After all buyers make their decisions, the experimental software displays the round results to the sellers. The way in which seller profit is calculated varies by treatment.

## 5. TREATMENTS

**Table 5. Our standard 2x3 experimental design**

| | Rank by relevance **(R)** | Rank by click-through **(CTR)** | Rank by relevance x click-through **(R x CTR)** |
|---|---|---|---|
| Pay per transaction **(PPT)** | Treatment 1<br>5 sessions<br>60 subjects | Treatment 3<br>5 sessions<br>60 subjects | Treatment 5<br>5 sessions<br>60 subjects |
| Pay per click **(PPC)** | Treatment 2<br>5 sessions<br>60 subjects | Treatment 4<br>5 sessions<br>60 subjects | Treatment 6<br>5 sessions<br>60 subjects |

The experiment has a standard 2x3 design. Sellers whose products are shown to the buyers are always charged the *Lowest Accepted Bid* in each product group (see Table 4). Depending on the treatment sellers are charged the Lowest Accepted Bid amount either per buyer click or per transaction. So total seller profit under PPT (pay per transaction) in a round is:

*Number_of_Units_Sold* x (*Price – Production_Cost – Lowest_Accepted_Bid)*

and total seller profit under PPC (pay per click) in a round is

*Number_of_Units_Sold* x (*Price – Production_Cost) – Number_Of_Clicks x Lowest_Accepted_Bid*

Note that the Lowest Accepted Bid, the Number of Units Sold, and the Number of Clicks can be different in Group 1, Group 2, and Group 3.

We also vary the order in which products are displayed on a buyer's screen. The display order is the second treatment variable. The search engine always picks to show the products with the highest 4 bids, however, the ordering of the products is different. The products are ranked and displayed by Relevance (R), by past Click-through Rate (CTR) or by the product of the two (R x CTR). Ties in rank or bids (for sellers) are broken at random.

***Example (con'd)****: Suppose that products are displayed in order of their relevance (R) to the buyers. For Buyer 1 the top four products to be shown are A1, B1, B2, A3 (see Table 4, column 2). These products will be displayed from top to bottom in the order of their value to Buyer 1, which is A1 first (value - 120), then A3 (value - 80), then B2 (value - 80), then B1 (value - 65) (See Table 2).*

Finally, each session consists of 30 rounds. Initially each round lasts 2 minutes (1 minute for sellers and 1 minute for buyers). As the experiment progresses subjects are able to make decisions faster, so the time for each round is reset to 1 minute and 20 seconds (40 seconds for sellers and 40 seconds for buyers). To prevent large differences in earnings, the production costs of sellers 4, 5, and 6 are switched in round 11 and round 21.

## 6. EXPERIMENTAL PROCEDURE

The experiment proceeds in the following fashion. First subjects are recruited from the undergraduate student population of a large urban business school. When they arrive at the laboratory, subjects are randomly assigned to be buyers or sellers in the experiment. Then subjects are assigned to a computer terminal and given computerized instructions. After reading the instructions subjects sign the informed consent form and participate in a practice round, which lasts 4 minutes. Then the actual experimental session starts. The experimental sessions lasts usually around 70-85 minutes. Subjects can ask questions any time during the experimental session. At the end subjects are paid in private the earnings that they received during the experimental session converted into US dollars plus a $10 show up fee. The average subject earnings were approximately $65.00.

## 7. RESULTS

All the results reported below are based on a standard two-factor ANOVA with replication. We have 5 statistically independent observations in each of the six cells of the experimental design (see Table 5). If the optimal allocation is implemented, the total surplus is 840 experimental dollars per round. The average efficiency attained during the experiment was 75% (or 630 experimental dollars per round).

### 7.1 Results related to total surplus

Our results show that the way in which products are ranked on buyers' screen has no effect on efficiency (p-value = 0.72). However, PPC definitely outperforms PPT (p-value = 0.018). Since the experimental environment is quite complex, we decided to probe further by comparing subject performance during the last 10 rounds of the experiment after subjects have had some experience with the experimental environment. We find that PPC still outperforms PPT (p-value = 0.008) but we can also detect a substantial difference due to ranking method. CTR outperforms R (p-value = 0.08) and R x CTR is somewhere in-between the two. If we break up the total surplus into its three components: buyer surplus, seller profits, and search engine revenues, we get some more interesting results.

### 7.2 Consumer surplus

Buyer surplus is not affected by the ranking method (p-value = 0.96) but seems to be higher under PPC than under PPT (p-value = 0.21). The p-value is relatively high but decreases if we take into consideration only the last ten rounds in each session (p-value = 0.048). In addition, we found that the variance of buyer surplus over the 30 rounds of the experiment is different under the treatments investigated. The variance is larger under PPT than under PPC (p-value = 0.016). There are also interaction effects, namely, the difference in variances of buyer surplus is lowest under CTR, higher under R, and even higher under R x CTR (p-value = 0.031). We detected no significant treatment differences in buyer search costs. We also found that the best product matches are shown most frequently to consumers in the CTR treatments although the best matches usually do not occupy the top slots. Consumers' clicks are mostly driven by their most relevant match (the products for which they have the highest value). That is why they click more often on the top two spots in the R treatments, but more often on the 3rd and 4th slot in the CTR treatments.

### 7.3 Advertiser profits

Seller profits are not affected by payment method (p-value = 0.70). This is interesting because under PPT sellers are clearly limited by their budgets and can receive a negative profit in extremely rare occasions. This happened only twice during all 15 sessions where sellers paid per transaction. On the other hand sellers quite frequently end up with a negative profit for the round in the PPC treatment because they are more optimistic about their conversion rate (this happened 431 times during the 15 sessions when sellers paid per click). Under these circumstances there is a revenue transfer from sellers to the search engine.

Curiously enough seller profits seem to be affected by ranking method (p-value = 0.21). The p-value is relatively high but there are additional clues that suggest this result might become more robust if more sessions are conducted. For example, again if we consider only the last 10 rounds in each session, the significance improves (p-value = 0.145). In addition, if we test only R vs CTR using all rounds, the significance is high (p-value = 0.06) and CTR outperforms R. Moreover, we found that the main driver behind this difference is the significantly higher prices of the product in the $4^{th}$ (cut-off) slot in CTR (p-value = 0.048). Generally it seems that sellers perform best under CTR, worst under R, and in-between under R x CTR. We also found that seller bids are lower under PPC (p-value =0.088) but this is expected since the number of clicks is almost always higher than the number of transactions. We did not detect any treatment differences in overall average product prices.

### 7.4 Search engine revenues

Search engine revenue is not affected by the ranking method (p-value = 0.88) but is affected by payment method. Search engine revenue is higher under PPC than under PPT (p-value=0.13). Search engine revenue is increasing steadily throughout the rounds in all treatments while buyer and seller surplus are either decreasing or relatively constant throughout the rounds most of the time (see Figure 1 in Appendix).

# 8. CONCLUSION

The main conclusion from our results so far is that the matching mechanism definitely performs best when advertisers (sellers) pay per click (PPC). Search engine revenues are higher under PPC because sellers are often too optimistic about their ad's conversion rates. Among themselves sellers are also relatively more expressive about their costs under PPC so for consumers this option is best because the best matches are displayed more often under PPC. There are some indications that the best ranking method would be CTR, since sellers prefer it somewhat. However, in order to confirm this result and probe more specifically into the underlying reasons for our main results, we need to extend the experimental design and conduct several sessions in each treatment with experienced subjects. Another interesting extension of this study will be to incorporate the Generalized Second Price rule as a matching mechanism and a relevance accuracy level less than 100% in an extended experimental design.

Overall it seems that Dellarocas and Viswanathan's model of the interaction between consumers, advertisers, and search engine most closely agrees with our results here, although the reasons for some of the effects we report are behavioral and not strictly rational by game-theoretic standards. We also need to carry out a deeper analysis of individual behavior to see which of the assumptions of the several theoretical models mentioned might need to be revised in order to improve predictions.

# 9. ACKNOWLEDGEMENTS

We thank Microsoft for providing funding for this research through the "Beyond Search" program

# 10. REFERENCES

[1] Athey, S. & Ellison, G. 2008 Position Auctions with Consumer Search, working paper, Harvard University and MIT, August 2008

[2] Blumrosen, L, Hartline, J., Nong, S. 2008 Position Auctions and Non-uniform Conversion Rates" Workshop on Ad Auctions '08 Chicago, Illinois USA

[3] Dellarocas, C., & Visawanathan, S. 2008 The Holy Grail of Advertising? Allocative Efficiency and Revenue Implications of 'Pay-per-Action' Advertising in Environments with Quality Uncertainty, working paper, University of Maryland, 2008

[4] Smith, V. L. 1976 Experimental Economics: Induced Value Theory. *American Economic Review*, Vol. 66, No. 2, *Papers and Proceedings of the Eighty-eighth Annual Meeting of the American Economic Association*, May 1976, pp. 274-279.

[5] Smith, V. L. 2003 Constructivist and Ecological Rationality in Economics. *American Economic Review*, Vol. 93 No. 3, June 2003, pp. 465-508.

[6] Thompson, D. and Leyton-Brown, K. 2008 Tractable Computational Methods for Finding Nash Equilibria of Perfect-Information Position Auctions, Workshop on Ad Auctions '08 Chicago, Illinois USA

[7] Vragov, R. 2009 Sponsored Search as a Strategic E-Service Journal of E-Services and Mobile Applications, v1 n1, January-March 2009

[8] Vragov, R. and Levine, I. 2007 A Pricing Algorithm for Mixtures of TV Programming and Commercials, IEEE Broadcast Technology Society Newsletter, Fall, 2007

# 11. APPENDIX

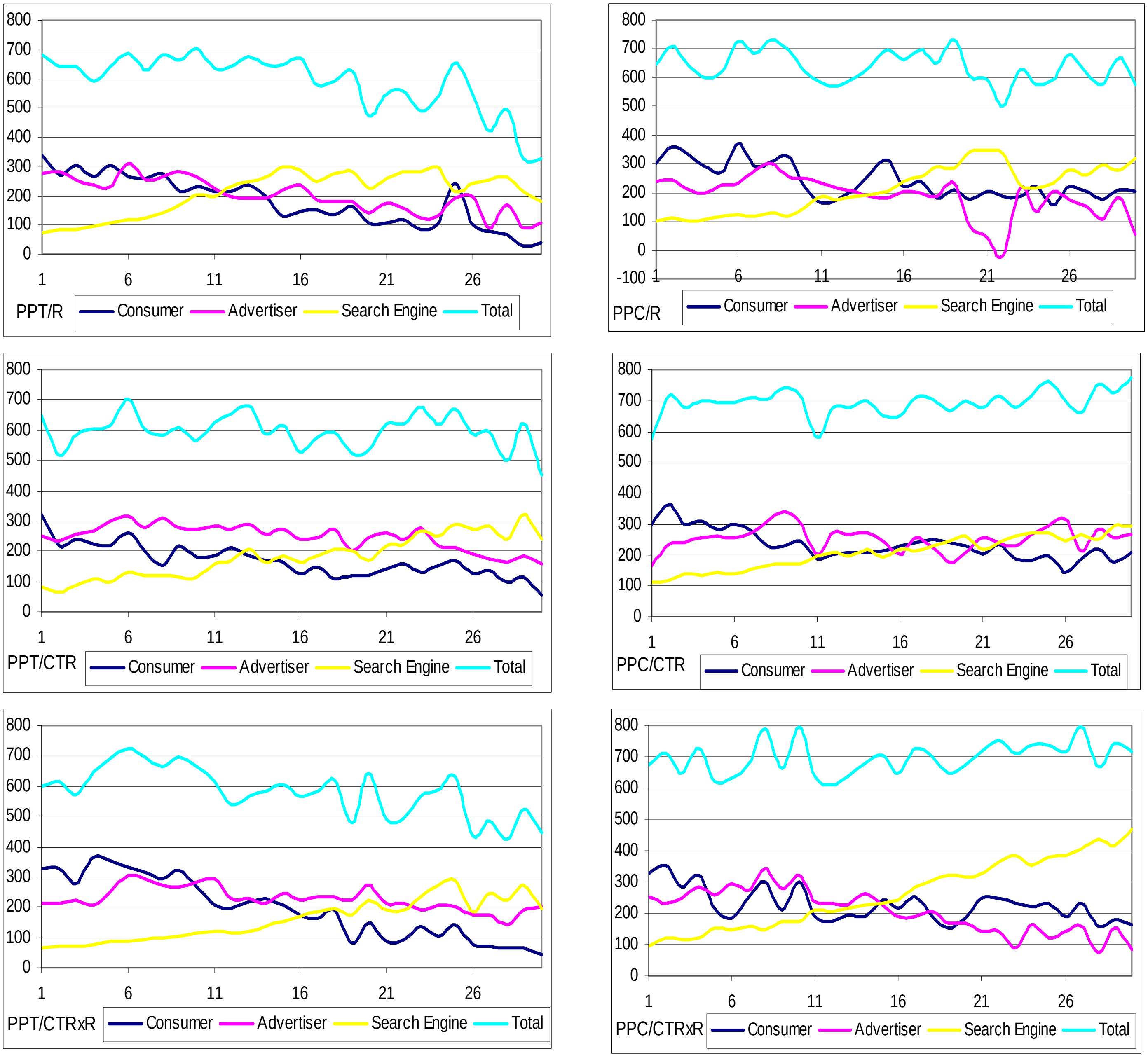


**Figure 1. Total surplus, consumer surplus, search engine revenues, and advertiser profits in each treatment of the experiment by round.**

# Exploring Collaborative Filtering for Sponsored Search

Sarah K Tyler
University of California, Santa Cruz
1156 High St.
Santa Cruz, CA 95064
skt@soe.ucsc.edu

Yi Zhang
University of California, Santa Cruz
1156 High St.
Santa Cruz, CA 95064
yiz@soe.ucsc.edu

Dou Shen
Microsoft
One Microsoft Way
Redmond, WA 98052
doushen@microsoft.com

## ABSTRACT

Sponsored search seeks to align paid advertisements with interested individual search engine users. Existing sponsored search algorithms are based on advertisers' bidding on individual search terms. Search terms, however, are not an accurate description of a user's information need or ads preferences. Additionally, advertisers are not always good at identifying all the search terms that are relevant to their products or services. On the other hand, collaborative recommendation systems assume users who have similar tastes on some items may also have similar preferences on other items, and thus make recommendations for one user based on the feedback from other similar users. In this paper, we explore whether collaborative filtering methods can help predict which users are likely to click on which advertisements. More specifically, we use the user supplied query as well as session based user information to build a better user profile for inferring the users' hidden information needs. We tried two basic collaborative filtering algorithms, a k-nearest neighbor, and a probabilistic factorization model, to determine whether collaborative filtering on sessions and queries can benefit sponsored search. The experimental results on 100 million Microsoft search impression data set demonstrates the effectiveness of the collaborative filtering approach.

## Categories and Subject Descriptors

B.3.3 [**Information Search and Retrieval**]: Information Filtering

## General Terms

Algorithms, Applications

## Keywords

Collaborative Filtering, Recommendation Systems, Sponsored Search



## 1. INTRODUCTION

Online Advertising generates the major part of the search engines' revenue and thus became one of the most important research problems among internet researchers. How to recommend advertisements that are relevant to a specific query issued by a specific user under a set of known circumstances is a challenging research problem. Even a small improvement in accuracy could lead to a big economic benefit for a company.

Despite the wide commercial and research interests in search advertising, the quality of ads recommended to search engine users still needs much improvement. The ad clickthrough rate of major search engines are much lower than the clickthrough rate of organic search results. Besides, 60%-70% of the query volume do not have any ads shown [6]. Optimizing for user experience and search engine profitability are not the same task [22]. Ad placement solutions that only use the biding price and biding query provided by advertisers are not optimal for search engine users. For example, most existing search engines return back advertisements that simply match a query. However, the real task is to satisfy the user's information or commercial needs. Existing systems assume that a user knows what words to use to describe his or her need. However, the user's information need is characterized by complex user criteria (e.g., relevance, price, trustworthy, etc.) and the context of the user, neither of which can be modeled easily given a few search terms. Similarly, the description of an ad is often minimal and not enough to infer which users/queries might be good matches for it.

For a list of advertisements presented to a user, the user will choose to click on none or a few ads. To some extent, this sponsored search scenario resembles the standard collaborative filtering (CF) task. Collaborative recommendation systems assume users who have similar tastes on some items may also have similar preferences on other items, and thus make recommendations for one user based on the feedback from other similar users. Typically, this is done by representing data as a user - item matrix with missing entries, and the task is converted into the task of predicting the missing entries. In sponsored search, the data can also be represented as a user (session or query) - ad matrix, where each entry corresponds to the clickthrough rate/count of the ad for that user. How to better infer the information needs of the user and improve user satisfaction are heavily studied in the collaborative filtering literature. These earlier works motivated us to explore whether we can adapt collaborative filtering techniques to improve sponsored search performance.

Sponsored Search also differs from the traditional collaborative filtering task. First, the two tasks have different basic elements. In traditional CF, there are two elements: users, items. However, in sponsored search, there are three elements: users, queries, ads. Unlike traditional CF, we cannot decide which ad is of interest purely based on the user, we have to also consider the queries. As a starting point to solve this problem, we use both "queries" and "sessions" to represent users. Second, the two tasks have very different distributions of data. In well studied traditional collaborative filtering domains, the average number of ratings per item is significantly larger than the average number of clicks per ads. For example, on the Netflix data set, there are 17 thousand products, 100,000,000 ratings and 500,000 users, which means an average of nearly 6 thousand ratings per movie.[20]. In contrast, the sponsored search data we studied contains close to 2 million ads with only 360 million instances for an average of 115 examples per ad. Thus Sponsored Search has a greater distribution in the long tail of ads with few examples. While it is important to maximize the click-through rate, it is also important to ensure that each advertisement generates some clicks, otherwise advertisers might go elsewhere. Thus while it may be tempting to only recommend ads for which there is ample training data, we cannot ignore the long tail of the distribution, which corresponds to ads with few training data. This is a well known cold start problem. Although it has been heavily studied for users, little work has been done to solve this problem for products.

Considering the similarity and difference between sponsored search and traditional collaborative filtering tasks, the questions is: whether standard collaborative filtering techniques can be used to improve sponsored search performance. The work described in this paper tries to answer this question. To do so, we use the implicit feedback of whether the user clicked on an ad to determine the effectiveness of the ad to the given user. We explore the traditional collaborative filtering techniques to the sponsored search task, Adword, to predict the click-through rate for each ad.

## 2. RELATED WORK

As the area of search engine advertising involves many "trade secrets" of commercial companies, the amount of publicly available related research papers is limited. [14] allow personalized advertising based on the user's location, browsing and interaction history. In the work of Ribeiro-Neto et al. [23], an impedance coupling technique is proposed for contextual ad placement. [18] proposes a system that is able to adapt online advertisement to a user's short-term interests. It relies on search keyword supplied by the user to search engines and on the URL of the page requested by the user. [19] uses a language model mixtures for contextual ad placement in personal blogs. [27] uses web search engine log files and various content features to extract keywords from web pages for advertisement targeting. Much work with online advertisements have focused on contextual advertisements. In this scenario an advertisement is placed in a web page given the content of the web page. Web pages often have static content, but can change quickly. Anagnostopoulos et al[2] proposed using summaries instead of full page content for selecting ads. They showed that the summation approach was often sufficient, and performed often as well as the full page analysis. Broder et al [7] used the full page, classifying it into one of 6000 categories. Das et al[8] used a two-tiered approach for classifying large collections of news data. The first approach used a relatively simple classifier to group similar pieces together. A more sophisticated classifier was then used on each subset of data individually. This is a similar approach to the one we are taking. [6] expanded queries to produce query rewrites for broad match based on relevance feedback techniques. As we will demonstrate later, the collaborative filtering technique also serves as a complementary query substitutions techniques for broader match.

## 3. COLLABORATIVE FILTERING

The basic assumption of collaborative filtering is that users that have similar preferences on some documents may also have similar preferences on other documents. Therefore the algorithm provides recommendations for a user based on the opinions of other like-minded users. Memory-based heuristics and model based approaches have been used in the collaborative filtering task [16] [9] [5] [15] [13] [11]. In collaborative filtering, user ratings over items are represented as a matrix $A$, where $A_{u,i}$ is user $u$'s rating on item $i$. Many collaborative filtering techniques have been proposed to predict the missing cells in the matrix [16, 9, 5, 15, 13, 11, 3, 1]. Sponsored search differs from collaborative filtering in that there are three elements: users, queries, ads. For simplicity, we can align user/query by either treating "queries" or "sessions" as the user, and then use collaborative filtering techniques for collaborative advertising. Without lose of generality, we introduce collaborative filtering techniques using traditional CF terminologies "user" and "item" in this section.

We first introduce two basic collaborative filtering techniques: K nearest neighbor, a probabilistic factorization approach. We choose the two techniques because they are commonly used, simple algorithms that are very different and complementary to each other. Variations of these techniques have been successfully used in the Netflix competition[17][3][4]. The goal of our work is to maximize the click through rate. There are different definitions of click through rate in the literature, and we define the clickthrough rate over a query-ad pair. Each instance is a (query, ad, impression) triple, where impressions means a page impression, which is generated every time a user views a page displayed by a search engine. In the rest of this paper, we use $q$ to represents a query, $a$ to represents an ad.

### 3.1 K-Nearest neighbors based on Item-item similarity or user-user similarity

There are two very commonly used collaborative filtering approaches: the first one compares each user to the other users, while the second one compares the items to each other. These are called *user-user similarity* and *item-item similarity* respectively. We describe Konstan et al.'s Pearson's correlation based user-user algorithm [16] below. For each user, we calculate the average rating assigned by that user $\bar{A}_u$ to all rated items. Each unknown rating is then estimated as the user's average rating, perturbed by the sum of the difference between every other user's assigned rating and his/her average rating, weighted by the correlation among the commonly rated items of the current user to every other user who are one of the $K$ nearest neighbors of the current user.

Formally, this is stated:

$$A_{u,i} = \bar{A}_u + \sum_{v=1}^{K} \frac{w_{u,v}(A_{v,i} - \bar{A}_v)}{|w_{u,v}|} \quad (1)$$

where $w_{u,v}$ is the Pearson's correlation between user $u$'s ratings and user $v$'s ratings. Recall that Pearson's correlation is:

$$\begin{aligned} w_{u,v} &= \frac{\sum_{j=1}^{m}((A_{v,j} - \bar{A}_v)(A_{u,j} - \bar{A}_u))}{\sigma_v \sigma_u} \\ &= \frac{\sum_{j=1}^{m}((A_{v,j} - \bar{A}_v)(A_{u,j} - \bar{A}_u))}{\sqrt{(\sum_{j=1}^{m}(A_{v,j} - \bar{A}_v)^2)(\sum_{j=1}^{m}(A_{u,j} - \bar{A}_u)^2)}} \end{aligned} \quad (2)$$

where $\sigma_u$ and $\sigma_v$ are the standard deviations in user $u$'s and user $v$'s ratings, and $m$ is the number items that user $u$ and user $v$ have both rated. Herlocker et al. [12] examined using other similarity methods such as Spearman's correlation, information entropy, mean-squared difference, and found they performed similar to Pearson's correlation. In its simplest form, the item-item algorithm is similar to the user-user algorithm, only with the rows and columns exchanged. Item-item similarity can be extended to take into account the content of the items being rated. For example, Sawar et al. [24] estimated ratings by summing the ratings of the other rated items, weighted by the cosine similarity of the rated and unrated plain text documents.

For our model we used a vector of ad click-through rates to represent each query. We first find the K nearest neighboring queries that have occurred with the ad. We compute query-query similarity by taking the cosine similarity of the vector ad click-through rates. Then, weighted average of the clickthrough rate of the ad from neighbors are used as the prediction of the current query-ad pair.

## 3.2 Probabilistic Factorization Model

The probabilistic factorization model we used is similar to regularized singular value decomposition. The major focus of this approach is also based on learning the hidden representations of each ad and query. We model the intention of a query as a vector $h_q$ to be learned from the data. We model an ad as a vector of $h_a$ to be learned from the data. We assume the prior distribution of $h_q$ follows a Gaussian distribution centered on the zero vector, and the prior distribution of $h_a$ follows a different Gaussian distribution centered on the zero vector.

We model the probability of an ad being clicked when displayed for a query as follows:

$$p(click|q,a) = c_q + c_a + h_q^T * h_a \quad (3)$$

where scalar $c_q$ is a query specific parameter, scalar $c_a$ is an advertisement specific parameter, and vector $h_q$ is a the hidden representation of the query and vector $h_a$ is the hidden representation of the advertisement.

We treat each (query, ad, impression, click) as a labeled training instance represented as a triplet $(c_i, q_i, a_i)$, where $c_i$ means whether ad $a_i$ is clicked when shown to query $q_i$, and i is the index to the instance. Then we can learn the parameters $\theta$=$(c_q, c_a, h_a, h_q)$ to minimize the error of predicting click through on the training data:

$\theta = argmin_{c_q,c_a,h_a,h_q} \prod_a P(h_a) \prod_q P(h_q) \prod_i P(c_i|q_i, a_i)$

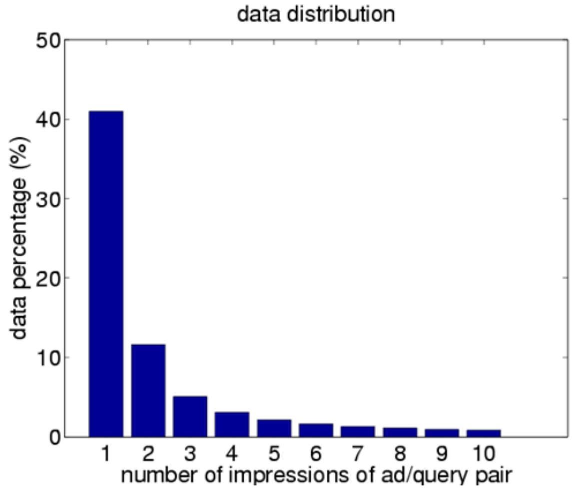


**Figure 1: The distribution of ad occurrences. The majority of ads occur rarely, with 40% of the ads occurring only once.**

# 4. EVALUATION METHODOLOGY AND EVALUATION DATA

One of the leading internet search engines, MSN, shared their online advertisements click-through data with the research community. The ad log data contains meta information about the ad: the time it was clicked, its relative position to other ads, as well as user information such as the session id of the user interacting with the search engine and the user supplied query. However the user IP address is not provided. For simplicity and as a starting point, we will only focus on straight forward collaborative filtering approaches, which ignore the additional meta information. The data set contained roughly 100 million impressions, about 360 million (query, ad, impression) triples, 1.8 million unique advertisements, 34 million unique sessions, and 28 million unique queries. 434 thousand (or 23%) of the ads occurred only once. Query-ad pairs occurred even less frequently. As seen in Figure 1, 40% of (query, ad, impression) triples are for query-ad pairs that occurred only once. Thus in order to do a good job of prediction on these query-ad pairs, it is important to be able to generalize well between users.

In order to perform collaborative filtering well we first need to identify which information could be used to model users. One traditional approach is to view the current user the same as the current query issued by the user, considering the fact that the query is the user's expression of his or her information need. However, query is inherently ambiguous and does not tell much about the intention of the user. Thus the approach of using queries as users may be an over simplification.

A second approach is using the information collected in the current session to model the user. The intuition is that sessions are a series of actions made by the user to achieve the user's current task. Often sessions include several queries issued closely together. Each query may be a refinement of the representation of the same underlying interest or goal. A session may also represent a user's changing information need. Given more user specific information, sessions may be

**Table 1: Example of similar sessions**

| $Session_1$ | $Session_2$ |
|---|---|
| CELL PHONE POWER | metro pcs deals |
| CELL PHONE RECEPTION POWER | free cell phones |
| CELL PHONE RECEPTION POWER COMPRASON | cell phone plan online comparison tool |
| Cell Phone Reception Comparison | compare alltel at&t suncom and surewest |

**Table 2: An example of the two most similar 5-query sessions in the training data as determined by KNN with cosine similarity. The intentions behind each query appear to be different, but each user is clicking on the same advertisements, which may indicate they have the same purchasing needs.**

more accurately represent users' hidden information needs. The major risk of this approach is straight forward collaborative filtering on sessions might not work since sessions are often very short, thus making it difficult to find good neighboring sessions given limited data.

## 4.1 Similar Sessions

It is worth noting how comparing similar sessions is influenced by the original sponsored search algorithm as well as user issued query. As with representing queries for query-query similarity we use a vector of ad clicks to represent each session. The value for each ad in the session is the percentage of click through a particular ad received during the session. Sessions that have similar ad click-through are similar. If two users in two separate sessions issue ambiguous queries, but click on the intersection of ads, they will appear to be similar despite the query ambiguity. Session similarity, however, is still dependent on the subset of ads presented to the user. The ads presented to the session are a result of the queries issued by the user. Thus two sessions that are similar, may have similar queries.

Consider the example shown in Table 1 which shows an example of session-session similarity. The queries intents are not the same. The user in the $session_1$, $user_1$ appears to be refining a query for the same task - comparing cell phone reception power. The user in $session_2$, $user_2$, also appears to be interested in cell phone comparisons, though it is not clear from just the queries whether reception power is of interest to the user. Since both users click on the same advertisements, we may be able to infer the brands $user_1$ is interested in, or that $user_2$ may also care about reception power.

# 5. EVALUATION

We evaluate the two collaborative filtering techniques over two tasks: reranking of advertisements and filtering advertisement results. Since collaborative filtering is best suited for helping rare user-ad pairs where user information is prevalent, we explore reranking primarily over rare data. We define rare and common data as follows.

- Rare query-ad pairs - $Queries, Ad$ tuples where the query and ad were seen together less than 10 times in our data.

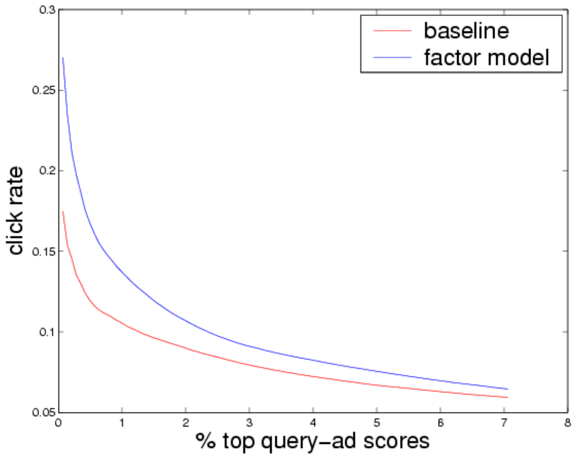


**Figure 2: P@K results for reranking of ads using matrix factorization.**

- Commonly issued queries - Queries that occured more than 10 times in our data. Similarly are queries were queries that occured less than 10 times in our data.
- Rare session-ad pairs - $Session, Ad$ tuples where the ad was seen in the session less than 10 times in our data.
- Long Sessions - Sessions where at least 10 advertisements were issued over all queries issued in the session. Similarly, Short Sessions are sessions that had less than 10 advertisements.

## 5.1 Mean Average Precision for ReRanking

In this experiment we re-rank the ads and evaluate the results with precision@k.

### *5.1.1 Matrix Factorization*

We carried out two experiments to understand the performance of the probabilistic factorization model.

In our First experiment, we used the last 90% of the click through data as training data, and tested the performance of the models on the rest 10% of the data. Based on the prediction of the models, we rank the query-ad pairs that never occurred in the training data and occurred once in testing data. We use a simple baseline as defined in Equation 4:

$$score = c_{query} + c_{ad} \tag{4}$$

where the score of a <query,ad pair> is a combination of the click through rate of the query, ($s_{query}$), and the click through rate of the ad, ($s_{ad}$).

Figure 2 shows that top ranking pairs have a very high average click through rate in the testing data. Modeling the hidden representation of the query and ads does provide improvement over the simple baseline. The evaluation is based on about 36 million testing points, and the improvement is significant.

In the second experiment, for each query, we re-rank the ads and compute the click through rate at top k (represented as %topk). Besides baseline described in Equation 4,

| | All Queries | | |
|---|---|---|---|
| | Baseline | Factorization | Original Ranking |
| %@top1 | 4.11 | 4.19 | 4.95 |
| %@top3 | 2.5 | 2.53 | 2.92 |
| %@top5 | 2.01 | 2.02 | 2.11 |
| | Queries Seen at Least 20 times | | |
| | Baseline | Factorization | Original Ranking |
| %@top1 | 5.76 | 5.9 | 6.58 |
| %@top3 | 3.14 | 3.17 | 3.58 |
| %@top5 | 2.4 | 2.41 | 2.5 |

**Table 3: Comparison of reranking for Factorization model over all queries and rare queries. %topK is the average clickthrough rate of the top K ads. The average clickthrough rate of all the query-ad pairs of the existing search engine is about 2%.**

**Table 4: Example of similar queries**

| Windows Defender | Irritable Bowel Syndrome |
|---|---|
| windows defender | irritable bowel syndrome |
| how do you flip through on microsoft vista | bacteria and IBS |
| clear qam on vista media center | ibs and prozac |
| Microsoft Windows Defender | bowel diseases |

**Table 5: An example of similar queries using KNN.**

we also use the original ranking of the advertisers provided by MSN as a second baseline. Table 3 gives the performance of the three algorithms. Averaging over all queries, the probabilistic factorization model comes close to achieving the click-through rate of the original ranking provided by MSN search engine. Although the model does not perform as well as the original ranking, we think the results are still interesting and promising. It is well known that the click-through rate is highly dependent on the position of the ads, and thus the original ranking creates a high bias for clicks. This bias favors the MSN search engine, as users are more likely to click the MSN's top ranked ads. Additionally, the search engine used much more information (more training data, information about the queries, ads, etc.) than the collaborative collaborative filtering approach, which only uses the clickthrough information provided. It's very likely that the collaborative filtering algorithm will complement existing search engine advertising algorithms, and achieve overall better performance.

### 5.1.2 *K-Nearest Neighbor*

We first look at some examples of query neighbors found by K-NN algorithm as shown in Table 5. The first row contains two original queries, and the following rows are the neighbors found. The list looks reasonable and the neighbors seem good candidates for query substitution in broader query match.

Next we examined the p@k for rare (less than 10) query-ad pairs for commonly issued queries, rare session-ad pairs for long sessions, rare queries and short sessions. The results are shown in Figure 3. We first observe that neither short sessions nor infrequent queries show much difference in

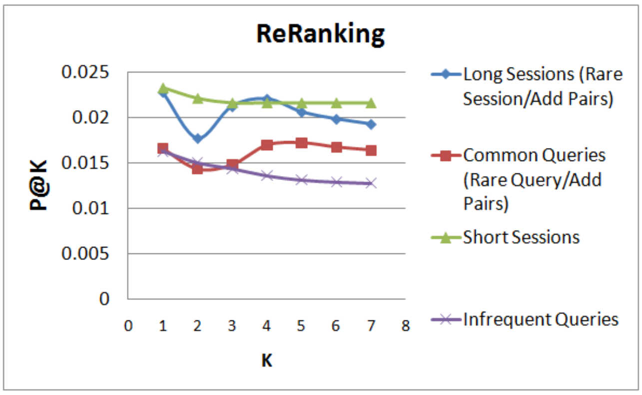


**Figure 3: P@K results for reranking of query-ad pairs using K Nearest Neighbor.**

p@k for different $k$. On the other hand, both long sessions (rare session/ad pairs) and common queries (rare query / ad pairs) appear to be more sensitive to $k$. This may be because rare sessions-ad pairs and rare query-ad pairs may be corresponding to rare ads.

The four curves show accuracy over four different datasets, so we cannot compare accuracy directly as one test set may be inherently more difficult than the others. While it is difficult to predict the user's intent from their query, we can easily determine that the click through rate which can give us some intution as to how hard the task is. If click throgh rate is high, then the search engine is better able to present ads that match the user's needs. Additionally, reranking of results with a large number of clicks becomes easier because there are more "good" ads to choose from. For common queries (rare query/ad pairs) and long session (rare session/ad) pairs, click through rate was identical at 1.9%, the same rate as the entire dataset. For common queries in general, however, the clickthrogh rate is 2.5% whereas the click through rate for long sessions is 0.7%. In fact, the top ten most common queries have a click through rate of 3.4% where as the top ten longest sessions have a click through rate of near 0. It seems the longer the session, the lower the click through rate. This may mean that the long session task may be more difficult and be the reason the short session curve appears to outperform the long session curve for $k = 2, 6, 7$.

## 5.2 Filtering of Search Results

In these experiments we use KNN as a classifier to predict an ad's click-through rate given the user data. We filter out the ads that were predicted to not be clicked. While not a traditional task for sponsored search, the task of filtering can provide additional means to understand the algorithms, since the filtering task is not as strongly impacted by the position bias we mentioned before. The task also has real world applications. For example, a mobile phone with a smaller screen cannot display the same amount of text legibility as a standard computer monitor. In this setting it may be desirable to prune lists of advertisements and remove ones less likely to be clicked.

We compute the mean precision by Equation 5. The precision is the number of click through ads not filtered over the number of ads not filtered.

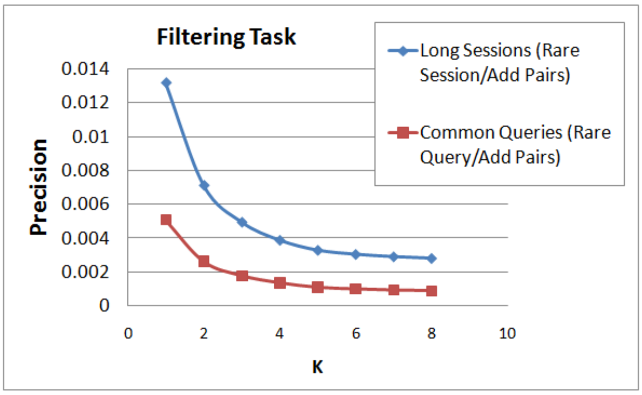


**Figure 4: Precision for the filtering task.**

$$precision(k) = \frac{\Sigma_{i=0}^{k} |U_{clicked}(i) \cap (U(i) - U_{filtered}(i))|}{\Sigma_{i=0}^{k} |(U(i) - U_{filtered}(i))|} \quad (5)$$

This precision metric is different from the traditional precision measures used in IR, since it is defined over an incomplete ranked list with missing entries filtering out. The motivation behind this precision metric is to minimize the original ranking bias. We assume all ads at position $k$ are effected by the ranking bias equally. If we filter the ads at position $k$ without changing the ranking of other neighboring ads (like what reranking does), we assume the ranking bias will have a minimal impact on our precision calculation. The results are shown in Figure 4.

As in section 5.1.2, the two curves show accuracy found over two different data sets. The curves suggest that long sessions are useful for filtering advertisements. Taken with the results of section 5.1.2, it seems long sessions may be as useful, if not more so than frequent queries. This may be because sessions are personalization and include the click-throughs of many user issued queries. More experiments are needed to identify the underlying reasons.

## 6. CONCLUSIONS AND FUTURE WORK

Can traditional collaborative filtering techniques be used to improve the sponsored search domain? The answer is yes. However, there are various ways one can apply CF to sponsored search. We have shown that collaborative filtering using queries as users is comparable to top search engine's existing approach, which uses much more information than CF algorithms. We also find that session information can be very useful for predicting click-through. One area for improvement would be combining these two different pieces of information about a user into a single CF model, thus adding personalization (session information) to global models (query information).

## 7. ACKNOWLEDGEMENTS

This research was supported in part by National Science Foundation IIS-0713111, a NSF graduate Fellowship, and a gift from Microsoft. Any opinions, findings, conclusions or recommendations expressed in this paper are the authors', and do not necessarily reflect those of the sponsors.

## 8. REFERENCES

[1] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17:734–749, 2005.

[2] A. Anagnostopoulos, A. Z. Broder, E. Gabrilovich, V. Josifovski, and L. Riedel. Just-in-time contextual advertising. *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 331–340, 2007.

[3] R. M. Bell and Y. Koren. Lessons from the netflix prize challenge. *SIGKDD Explor. Newsl.*, 9(2):75–79, 2007.

[4] R. M. Bell and Y. Koren. Lessons from the netflix prize challenge. *ACM SIGKDD Explorations Newsletter*, 9(2):75–79, December 2007.

[5] J. S. Breese, D. Heckerman, and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. Technical report, Microsoft Research, One Microsoft Way, Redmond, WA 98052, 1998.

[6] A. Broder, P. Ciccolo, E. Gabrilovich, V. Josifovski, D. Metzler, L. Riedel, and J. Yuan. Online expansion of rare queries for sponsored search. In *In Proceedings of the World Wide Web Conference*, 2009.

[7] A. Broder, M. Fontoura, V. Josifovski, and L. Riedel. A semantic approach to contextual advertising. *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 559 – 566, 2006.

[8] A. S. Das, M. Datar, A. Garg, and S. Rajaram. Google news personalization: Scalable online collaborative filtering. *Proceedings of the 16th international conference on World Wide Web Conference*, pages 271 – 280, 2007.

[9] J. Delgado and N. Ishii. Memory-based weightedmajority prediction for recommender systems. In *ACM SIGIR'99 Workshop on Recommender Systems*, 1999.

[10] S. Funk. Try this at home. http://sifter.org/ simon/journal/20061211.html, December 2006.

[11] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl. An algorithmic framework for performing collaborative filtering. In *SIGIR '99: Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 230–237, New York, NY, USA, 1999. ACM Press.

[12] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl. An algorithmic framework for performing collaborative filtering. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 99)*, pages 230–237, New York, NY, USA, 1999. ACM.

[13] T. Hofmann and J. Puzicha. Latent class models for collaborative filtering. In *IJCAI '99: Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, pages 688–693, San Francisco, CA, USA, 1999. Morgan Kaufmann Publishers Inc.

[14] D. Inc. Dart advertisement system.

http://www.doubleclick.com.
[15] R. Jin, J. Y. Chai, and L. Si. An automatic weighting scheme for collaborative filtering. In *SIGIR '04: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 337–344, New York, NY, USA, 2004. ACM Press.
[16] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl. GroupLens: Applying collaborative filtering to Usenet news. *Communications of the ACM*, 40(3):77–87, 1997.
[17] Y. Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *ACM Int. Conference on Knowledge Discovery and Data Mining (KDD'08)*, 2008.
[18] M. Langheinrich, A. Nakamura, N. Abe, T. Kamba, and Y. Koseki. Unintrusive customization technqiues for web advertising. In *Comput. Networks 31*, 1999.
[19] M. G. Mishne and M. de Rijke. Language model mixtures for contextual ad palcement in personal blogs. In *5th International Conference on Natural Language Processing*, 2006.
[20] Netflix. Netflix prize. http://www.netflixprize.com (visited on Nov. 30, 2006), 2006.
[21] A. Paterek. Improving regularized singular value decomposition for collaborative filtering. In *Proceedings of the KDD Cup and Workshop at the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2007.
[22] F. Radlinski, A. Broader, P. Ciccolo, E. Gabrilovich, V. Josifovski, and L. Riedel. Optimizing relevance and revenue in ad seach: A query substitution approach. *Proceedings of the 32th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 559 – 566, 2008.
[23] B. Ribeiro-Neto, M. Cristo, P. Golgher, and E. de Moura. Impedance coupling in content-targeted advertising. In *SIGIR*, 2005.
[24] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on the World Wide Web (WWW '01)*, pages 285–295, 2001.
[25] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl. Incremental singular value decomposition algorithms for highly scalable recommender systems. In *Proceedings of the 5th International Conferenceon Computer and Information Technology (ICCIT 2002)*, 2002.
[26] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. T. Riedl. Application of dimensionality reduction in recommender system – a case study. In *Proceedings of the Workshop on Web Mining and E-Commerce Systems at KDD 2000 (WEBKDD 2000)*, August 2000.
[27] W. Tau-Wih, J. Goodman, and V. Carvalho. Finding advertising keywords on web pages. In *In Proceedings of the World Wide Web Conference*, 2006.

# Towards Advertising on Social Networks

Maryam Karimzadehgan
Department of Computer Science
University of Illinois at Urbana-Champaign
mkarimz2@uiuc.edu

Manish Agrawal
Department of Computer Science
University of Illinois at Urbana-Champaign
magraw@microsoft.com

ChengXiang Zhai
Department of Computer Science
University of Illinois at Urbana-Champaign
czhai@cs.uiuc.edu

## ABSTRACT

Web advertising has become a financial backbone of business success nowadays. All major Web search engines such as Google, Microsoft and Yahoo! derive significant revenue from advertising. However, as a new area of research, online advertising has not yet reached its full potential. In particular, little research has been done on advertising on social networks. In this position paper, we present our review of some research issues related to advertising on social networks and some preliminary results in a related task of recommending news articles to users of Facebook.

## Categories and Subject Descriptors

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

## Keywords

Algorithms, Advertising, User Modeling, Social Network

## 1. INTRODUCTION

Nowadays, the Web has become an integral part of our lives. The prevailing business model of Web search relies heavily on advertising. A major part of advertising is *textual ads* which are short textual messages. There are two types of advertising: (1) *Sponsored search* places ads on the result pages from a web search engine according to the user's query. All major web search engines support such ads. Usually an ad consists of a title (3-5 words long), a description (around 20 words) and a URL that users are directed to by clicking on the ad. (2) *Content Match or Contextual Advertising* displays ads within the content of third-party Web pages.

As an emerging research area, online advertising has attracted much attention recently. The previous work can be summarized briefly as follows:

**Ads matching:** A lot of previous works have focused on developing methods to match pages to ads [20, 8, 4]. All these methods extract some features related to web pages to relate ads to pages. Some other works reduce the contextual advertising problem to sponsored search by extracting phrases from pages and matching those phrases with the bid phrase of ads [24].

**Query expansion:** Since user queries are short, some other works [18, 8] use additional sources of information for ad selection to expand users' short queries. In this approach, offline query rewriting is done by using various sources of external information and thus can only be applied to repeating queries. In a more recent work [7], authors propose a more efficient online expansion-based algorithm. Their algorithm builds an expanded query by leveraging offline processing which is done for related popular queries. Their results show the effectiveness of such a method for advertising on rare queries.

**Clickthrough prediction:** In online advertising, predicting the clickthrough rates; i.e., the number of clicks a given ad will solicit if it is displayed on the Web page is done previously. Authors in [19, 21] predict clickthoughs by clustering ads by their bid phrases and by analyzing the different parts of the ads (e.g., bid phrase and title, . . . ), respectively. However, these works focus on ad-based features to predict the clickthrough for a new ad. Authors in [5] study the intention underlying users' queries. They showed that clickthrough features such as deliberation time are effective in detecting query intent.

It seems that most research in online advertising has been focused on improving the *relevance* of the displayed ads to the page content. In other words, all these methods focus on maximizing the *match* between individual ads and the content of the page. However, there are other factors which also play an important role in effective advertising. In particular, as in the case of search, accurate understanding of a user's interest and need is critical for effective advertising. A lot of previous works [1, 13, 23] of modeling the behavior of Web search engine users have shown improvement in ranking documents by Web search engines. It is thus important to study how to improve user modeling for advertising.

In this position paper, we suggest that tapping into the growing research on social networks opens up many interesting opportunities to obtain more knowledge about users, thus potentially improving the effectiveness of online advertising. Compared with the traditional sponsored search and contextual advertising, advertising on social networks has not been studied much yet. The purpose of this paper is to lay out some interesting research issues related to advertising on social networks and to discuss some preliminary results from a related task to advertising on social networks


SIGIR-IRA, July 2009, Boston, USA.

– recommending news articles to users of Facebook.

# 2. ADVERTISING ON SOCIAL NETWORKS

People usually live in some communities and are associated with (potentially multiple) social networks. A human social network can be a group of friends living within a city, or a group of college classmates who remain in frequent contact. It can also be a group formed specifically to accomplish a set of tasks over time. Social networks are well-trusted because of shared experiences and the perception of shared values or shared needs [10]. For example, friends tell friends about restaurants and movies. These characteristics of social networks have two important potential benefits for advertising:

**First**, advertising propagated through a social network can be expected to be more trustable. Indeed, people in social networks are often more willing to trust and accept recommendations from their neighbors. It is human nature to be interested in what a friend buys more than in what an anonymous person buys, to be more likely to trust a friend's opinion, and to be more influenced by a friend's actions. A Lucid Marketing survey found that 68% of individuals consult friends and relatives before purchasing home electronics – more than the half who used search engines to find product information [9].

**Second**, social networks potentially allow us to obtain valuable information about users through observing their activities. Moreover, social communities of users can also be leveraged to infer a single user's interest in the same spirit as collaborative filtering. All these indicate that we can potentially leverage social networks (particularly interactions and relations of people) to better model users and improve effectiveness of advertising.

Based on this analysis, we believe that the three most interesting high-level research issues about advertising on social networks are:

- *Advertising via relations*: How to effectively advertise through relations and interactions of people in a social network?
- *User modeling based on social networks*: How to leverage social networks to obtain an accurate model of user interests and needs?
- *Evaluation*: How to evaluate the effectiveness of an advertising system in a social network?

## 2.1 Advertising via relations

Unlike conventional Web search, if query terms match some documents in the index, this query will lead to some results, whereas in Web advertising, if no good results are available for the user query, it is better/desirable not to show any ad results. In other words, showing irrelevant ads would annoy the user and not yield any economic benefits [6]. Indeed, a study in [22] confirms that ads need to be relevant to the user's interest to avoid impairing the user's experience.

Social networks offer unique opportunities for advertising through relations and interactions of people, which can increase the trust of users in the advertisement. Patterns of influence and cascading behaviors have been studied in social networks [17, 16, 12] previously. In [16], authors have studied a very large recommendation network and observed the propagation of recommendations in such a large social network. Their findings show that the recommendation chain does not grow very large and it terminates after the initial purchase of the product. They also observed that the product will propagate through a very active social network. In addition, they defined the stochastic model which explains the propagation of recommendations. Authors in [12] have also measured the network value of a customer. For each customer, they model the probability of buying a product as a function of both the intrinsic properties of the customer and the product and the influence of the customer's neighbors in the network. Indeed, such studies open up new research directions and challenges in social networks for advertising.

The following are some interesting additional challenges for advertising through social networks:

- All social networks evolve over the time; as a result the advertising algorithm should take into account such evolutions, making modeling network evolutions an interesting challenge. If the network and this evolutionary behavior are well understood, it may be possible to drive a network to a profitable state.
- Modeling the influential nodes in social networks according to time for advertising is another challenge. These influential nodes are good targets for advertising as they could also influence others.
- Choosing an optimal set of users (group of people) to send advertisement to so as to maximize network profit is an NP-hard problem. Approximating algorithms will need to be developed to incorporate multiple factors such as relevance of an advertisement to a user, influentiality of users [11], and potential profit of an advertisement
- Modeling how user's interactions and interests would change not only with time but also with the specifications of the product is yet another challenge. These are related to the user's interests which might also evolve over time. Modeling the evolution of interests of users in social networks is very challenging.

## 2.2 User Modeling in Social Networks

In social networks, we can gain more knowledge about a user, but the integration of this wealth of information also presents challenges. In general, users in social network for advertising can be modeled in two ways:

- Gaining direct knowledge about the users. For example, in Facebook, we can gain information about the networks/groups one belongs to, activities one does, one's friends' networks and a wealth of information one can post on his profile such as links to the news, youtube links and a lot more.
- Inferring the behavior and preferences of a user based on knowledge about other people on a social network that the user interacts with.

A major challenge here is how to integrate these pieces of information.

In addition to characterizing a user based on the information associated with the user in a social network, we may further incorporate other relevant external sources such as user's blog posts, query logs, homepages etc., leading to an even more challenging question of how to gain all such knowledge and then integrate it.

## 2.3 Evaluation

In general, evaluation of computational advertising may involve multiple performance factors (e.g., profit of product providers, profit from placing ads, and user experiences). Depending on which factor(s) to emphasize, we may need different evaluation methods. When relevance of advertisement is the primary factor for evaluation, we may adapt existing evaluation methods for information retrieval to evaluate advertising on social networks. However, since ad relevance is much more subjective than topic relevance, creating a static gold standard test collection may be difficult, making it a significant challenge to directly adopt the standard Cranfield evaluation methodology.

A more promising solution may be to use the logs of ad clicks to quantitatively evaluate an advertising algorithm by assuming a clicked ad to be relevant and a skipped one nonrelevant [15]. Clearly, this evaluation strategy requires the deployment of a prototype advertising system and careful logging of user activities. There is also the challenge in developing an optimal interleaving strategy to compare different advertising algorithms.

# 3. LESSONS FROM A FACEBOOK NEWS RECOMMENDATION SYSTEM

As a study of user responses to recommending information through social networks, which is related to advertising on social networks, we present some lessons learned from developing a Facebook application for recommending news articles (called *Facebook Newsletters*). Since recommendation of news to users of a social network resembles advertising on social networks, some observations with our system may shed some light on the promise of advertising on social networks. Preliminary results indicate that most users find such an application useful and easy to use. It also shows that users of Facebook welcome recommendations given by their friends.

## 3.1 Overview of Facebook Newsletters

Facebook is one of the fastest growing social networks. It consists of many networks, each based around a school, workplace, or a region. It has users ranging from college students to working professionals. More than 100 million users log on to Facebook at least once each day [14]. As an experimental system for recommending information over social networks, we developed *Facebook Newsletters*, which provides daily newsletters for communities on Facebook.

In Facebook, each user has a personal profile and most users belong to one or more networks. Also, a user may join various interest-based groups on Facebook or may even start a new group. So, two users may be connected in three different ways: 1) User $a$ is a friend of User $b$. 2) User $a$ and User $b$ belong to the same network. 3) User $a$ and User $b$ belong to the same group. We refer to both network and group as "community".

A user may register a community by providing a keyword description and a set of news sources. The system then fetches the news articles from the specified sources (as well as standard sources such as Yahoo! news), and filters them based on the community description to prepare the daily news digest. It also prepares a list of popular articles for each community based on user feedback by using collaborative filtering techniques.

The application uses Facebook API to find out the networks of the user. The newsletters for the networks are available as *tabs* on the top of the page. The newsletter is presented as a list of articles, each with its title, news synopsis, links to original article and the locally cached page. Users can rate an article on a scale of 1 to 5. In the newsletter, the news articles are clustered, and the clusters are sorted on the score of the most relevant document in each cluster. Only one result per cluster is presented to the user, but the user can look at the other results by navigating through the *"Similar pages"* link. Only the top 5 results are presented to the user. The users can also *recommend* particular news articles to their friends through a recommendation button provided beside each news result. When a user makes a recommendation, his friend is sent a notification.

The recommendation of news articles was initially based on matching news articles with manually created keyword descriptions of a community. Matching articles are clustered using a centroid-based agglomerative clustering algorithm to alleviate the problem of redundancy. After obtaining user feedback, the system would improve recommendation decisions based on feedback information. The system is designed to collect and leverage the following three kinds of feedback information: (1) Clickthroughs of articles; (2) Ratings of articles; and (3) Recommendation of an article by a user to another user. All the feedback information is combined heuristically to improve filtering accuracy. More details about the system can be found in [2, 3].

| Statistics for all the communities: | |
|---|---|
| No. of people registered: | 60 |
| No. of clicks: | 350 |
| No. of recommendations: | 15 |
| Average rating (out of 586 ratings): | 3.41 |
| Average rating for clicked article: | 3.71 |
| Average rating for recommended articles: | 4.07 |

**Table 1: Usage Statistics**

## 3.2 Results of a pilot study

The *Newsletters* application was launched on Facebook for three months to conduct a pilot study of user responses. The application was advertised amongst university students by "word of mouth" publicity. In about three months, sixty people have added the application on Facebook from a number of universities. There were initially 3 seed communities on the application with available newsletters but for the period of this study (three months), this number has increased to 25. All the new communities are user initiated.

Table 1 shows the basic statistics of user clickthroughs and ratings. We can see that the average rating of a news article is 3.41 (out of 5), and it is higher for clicked and recommended articles. From these results, it seems that the articles that users clicked or recommended are the best candidates for including in the user feedback. This result suggests that if a user finds an article interesting, he would recommend the article to those friends who might also have an interest. This observation is true for advertising for *products* in that if a user finds a product useful, he suggests that product to those friends of him who have also indicated interest in that product.

A set of University of Illinois students were asked to use the application on a *regular basis*. They used the application for about one month and then a user survey was conducted to garner explicit feedback from them. Twenty two users participated in the survey.

The survey results indicate that most users plan to continue using the application in future and said that they at least got one *interesting* article every time they used the application. 18 users (81.8%) said that they got some articles that they *would not have gotten otherwise*, through their newspaper or regular web browsing. This is a very encouraging result. 95% of the users found the application from somewhat useful to very useful. 95% found it fairly easy to use. Only 14% think that the application helped them to socialize with their friends but another 77% feel that the application can potentially do that. Socializing would help a user to have more friends, as a result help them to receive more useful recommendations from them.

Overall, our findings clearly suggest that users in a popular social network such as Facebook are generally willing to accept recommendations of information through a recommender system that leverages user communities in social networks. While we should be cautious in generalizing the findings here to the context of advertising as there is clearly difference between recommending news and products, it is reasonable to hypothesize that advertising on social networks based on a similar recommender system may also be acceptable to users and users may receive interesting products through recommendation (from either their friends or the system) that may not be easily found through web search or browsing.

# 4. SUMMARY

In this position paper, we discussed why advertising on social networks is promising and presented some major research questions related to advertising on social networks. We also reported some preliminary results of a user study on a popular social network; Facebook, to understand the feasibility of recommending news to users in a social network. The results indicate that users find such a recommender system acceptable and useful, suggesting that it is potentially feasible to deploy a similar recommendation system on social networks for advertising.

# 5. REFERENCES


[1] E. Agichtein, E. Brill, S. Dumais, and R. Rango. Learning user interaction models for predicting web search result preferences. *SIGIR*, pages 3–10, 2006.

[2] M. Agrawal. Newsletters: An online news recommender system for social networks, 2008. M.S. Thesis, UIUC.

[3] M. Agrawal, M. Karimzadehgan, and C. Zhai. An online news recommender system for social networks. *SIGIR-SSM*, 2009.

[4] A. Anagnostopoulos, A. Z. Broder, E. Gabrilovich, V. Josifovski, and L. Riedel. Just-in-time contextual advertising. *CIKM*, pages 331–340, 2007.

[5] A. Ashkan, C. Clarke, E. Agichtein, and Q. Guo. Characterizing query intent from sponsored search clickthrough data. *SIGIR-IRA*, pages 15–22, 2008.

[6] A. Broder, M. Ciaramita, M. Fontoura, E. Gabrilovich, V. Josifovski, D. Metzler, V. Murdock, and V. Plachouras. To swing or not to swing: Learning when (not) to advertise. *CIKM*, pages 1003–1012, 2008.

[7] A. Broder, P. Ciccolo, E. Gabrilovich, V. Josifovski, D. Metzler, L. Riedel, and J. Yuan. Online expansion of rare queries for sponsored search. *WWW*, pages 511–520, 2009.

[8] A. Z. Broder, M. Fontoura, V. Josifovski, and L. Riedel. A semantic approach to contextual advertising. *SIGIR*, pages 559–566, 2007.

[9] K. Burke. As consumer attitudes shift, so must marketing strategies. 2003.

[10] E. K. Clemons. The future of advertising and the value of social networks. *Wharton ISE Blog*, pages 1–16, 2007.

[11] J. K. D. Kempe and E. Tardos. Maximizing the spread of influence in a social network. *SIGKDD*, pages 137–146, 2003.

[12] P. Domingos and M. Richardson. Mining the network value of customers. *SIGKDD*, pages 57–66, 2001.

[13] D. Downey, S. Dumais, D. Liebling, and E. Horvitz. Understanding the relationship between searchers' queries and information goals. *CIKM*, pages 449–458, 2008.

[14] Facebook. Facebook statistics. *http://www.facebook.com/press/info.php*, 2009.

[15] T. Joachims. Unbiased evaluation of retrieval quality using clickthrough data. In *SIGIR Workshop on Mathematical/Formal Methods in Information Retrieval*, 2002.

[16] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. *EC*, pages 228–237, 2006.

[17] J. Leskovec, A. Singh, and J. Kleinberg. Patterns of influence in a recommendation network. *PAKDD*, pages 380–389, 2006.

[18] F. Radlinski, A. Broder, P. Ciccolo, E. Gabrilovich, V. Josifovski, and L. Riedel. Optimizing relevance and revenue in ad search: A query substitution approach. *SIGIR*, pages 403–410, 2008.

[19] M. Regelson and D. Fian. Predicting click-through rate using keyword clusters. *Workshop on Sponsored Search Auctions*, pages 408–421, 2006.

[20] B. Riberio-Neto, M. Cristo, P. B. Golgher, and E. S. de Moura. Impedance coupling in content-targeted advertising. *SIGIR*, pages 496–503, 2005.

[21] M. Richardson, E. Dominowska, and R. Rango. Predicting clicks: Estimating the click-through rate for new ads. *WWW*, pages 521–530, 2007.

[22] C. Wang, P. Zhang, R. Choi, and M. D. Eredita. Understanding consumers attitude toward advertising. *8th Americas Conference on Information Systems*, pages 1143–1148, 2002.

[23] R. W. White, M. Bilenko, and S. Cucerzan. Studying the use of popular destinations to enhance web search interaction. *SIGIR*, pages 159–166, 2007.

[24] W. Yih, J. Goodman, and V. R. Carvalho. Finding advertising keywords on web pages. *WWW*, pages 213–222, 2006.

# A Preliminary Study on Dynamic Keyword Extraction for Contextual Advertising

Wen Ye[1,2], Wenjie Li[1], Furu Wei[1], Chunbao Li[2]

Department of Computing
The Hong Kong Polytechnic University, Hong Kong
{cswenye, cswjli, csfwei}@comp.polyu.edu.hk

Computer School
Wuhan University, China
licb1964@126.com

## ABSTRACT
Traditional ads keyword extraction approaches process a Web page as a whole. However, many current Web pages like Weblogs and discussion forums allow people to leave their comments, responses or follow-up questions on popular topics. Due to interaction among active participants, these pages often exhibit different focused topics in different places on the pages. In this paper, we emphasize on the linking relations that are built upon replies and quotations and propose a novel dynamic extraction approach for both inter-post and whole page ads keyword extraction. Preliminary evaluation results on Chinese forum data set demonstrate the effectiveness of the proposed approach.

## Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

## General Terms: Algorithms, Economics

## Keywords
Contextual advertising, dynamic keyword extraction

## 1. INTRODUCTION
The explosive growth of the Internet as a publication and interactive communication platform has created an electronic environment that is changing the way business is transacted. Nowadays, the Internet has become an important advertising venue. A large portion of the advertising market over the Internet consists of textual advertisements or ads, which encompass short text messages distributed to the users. Two main channels used to distribute textual ads are (1) sponsored search advertising that is driven by user's search query, and (2) contextual advertising that is driven by the content of a Web page. Unlike sponsored search advertising, contextual advertising requires a process of extracting advertising keywords from a Web page (i.e. keyword extraction) and a process of matching the extracted keywords against advertisements in an ads database (i.e. ads matching). The work presented in this paper is concerned with the problem of keyword extraction for contextual advertising.

Traditional keyword extraction approaches identify a set of keywords from a Web page by analyzing the content of it as a whole. These approaches work well on relatively stable pages like news pages, which do not change over time frequently. However, many current Web pages like Weblogs and discussion forums are much more strongly interactive among active participants. They allow people to leave their comments, responses or follow-up questions etc. on popular topics. The content of such kind of topic pages is assembled dynamically when new posts and new comments come forth. From time to time, the new posts on a topic page are likely to deviate from the initial topics. This is true especially when the topics are most discussed and the number of the people involved increases rapidly. It is thus reasonable to insert different ads in different places on a Web page when the topics on the page evolve or shift. Even in the same place, the ads to be attached may also need to be updated in response to post content on the increase. An approach to dynamic keyword extraction is desired.

In this study, we emphasize on the linking relations built upon people's replies and quotations that are prevalent in Weblogs and discussion forums. With linking information, we incorporate the influence from the related posts on a topic page to identify more suitable ads keywords in a particular position. Besides, we also use this information to update the existing ads arrangement when posts grow as time goes on. The remainder of this paper is organized as follows. Section 2 briefly reviews related work. Problem statement and formulation is given in Section 3. Then, a novel dynamic keyword extraction approach is introduced in Section 4. Following preliminary evaluation and discussion presented in Section 5, the paper is concluded in Section 6.

## 2. RELATED WORK
Early approaches to advertising keyword extraction identified keywords from a page based on the traditional tf*idf weighting method. Basically, words or phrases appeared much more times in a page have higher weighting scores. The ads relevance was then estimated according to co-occurrence of the same keywords within an ad and a page [1]. However, the matching mechanism based solely on the keywords identified from the text of a page could lead to many irrelevant ads. For example, if the keyword "apple" was identified, shall the ad about "apple pie" or the ad about "apple iPod" is delivered?

In order to overcome this problem, Broder et al. proposed an enhanced matching mechanism that combined a semantic phase with the traditional keyword matching phase. The semantic phase classified pages and ads depending on a pre-defined topic taxonomy and used ad and page classes as a factor in ad relevance estimation [2]. As they claimed, the approach worked quite well on the pages which involved only one topic. Notice that in their work the keywords were extracted for the general ads that were normally displayed around the main content area, such as the banner/footer ads at the top/bottom of a page and the ads in the left/right sidebars. The ads in these places were supposed to be relevant to the whole page.

Based on the analysis of Weblog comments, which was conducted by Misne et al. [3], Hu et al. introduced an approach to comment-oriented blog summarization [4]. The representative sentences were extracted from the post using the information hidden in its comments so that the extracted sentences could represent the topics that were concerned by the people who delivered the comments. The work presented in this paper is related but different from Hu's summarization work. While they employed the weights of the words appearing in comments to evaluate blog sentences, we explore explicit and implicit relations among the posts on forum topic pages to generate time-dependent keywords for targeting potential ads in various places of a page.

## 3. PROBLEM STATEMENT AND FORMULATION

Given a topic page, it is preferable to have ads closely relevant to its content in order to provide a better user experience and thus to increase the probability of clicks and to earn more profit. It has been recognized in the past that if an ad is displayed in a place between two posts, it will be noticed by more people. As a matter of fact, many sites have already done so. For example, Tianya (www.tianya.cn), one of the most popular forums in China, chooses to insert three ads between the 1st and the 2nd posts. Meanwhile, Discuz, the leading Chinese forum solution provider, attaches ads to all the posts in its supporting forums (http://www.discuz.net).

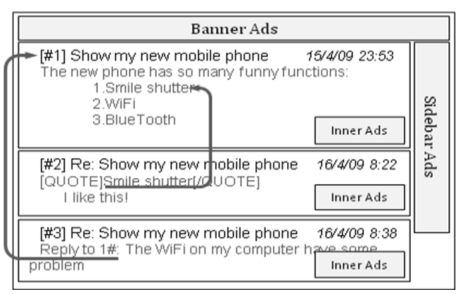


**Figure 1. A Forum Topic Page**

In this work, we consider two types of ads. One is the banner/sidebar ads, which are common in advertising and are selected based on the content of the whole page (as illustrated in Figure 1). The keywords extracted for this type of ads are called global keywords. The other is the inner ads which are to be placed between two posts and are mainly relevant to the content of the surrounding posts. The keywords extracted for this type of ads are called inter-post keywords. Both global and inter-post keywords are extracted dynamically. In particular, once a new post is uploaded, the global keywords may be changed due to word re-distribution. Not only that, the inter-post keywords of the existing posts may also get a chance to be updated due to the linking relations among the posts, such as the post #3's response to post #1 and the post #2's quotation from post #1 in Figure 1.

Currently, we simply consider the keywords between all any pair of posts for potential advertising. However, one may choose some appropriate places to insert ads, rather than putting ads everywhere on a page. This is an important and interesting issue but it is beyond the scope of the work presented in this paper. Now, let's define the aforesaid problem formally. Assume a topic page contains $n$ posts, i.e. $P=\{p_1, p_2, \ldots, p_n\}$. Each post $p_i$ ($i$=0, 1, …, $n$) embodies a posting time $t(p_i)$ (or simply $t_i$), an optional subject $s_i$, a main content body $b_i$. Each post $p_i$ is associated with an optional reply session $r_i$={the post being replied $pr_i$, the quotation $q_i$}, where the quotation may include either the whole or the partial content of the post being replied. In the case where a post reply session does not contain any explicit quotation, we assume that the whole content of the post being replied are implicitly quoted.

Our objectives are to extract maximum $m$ keywords for the banner/sidebar ads and for the inner ads inserted under each post $p_i$ at a given time point $t$, i.e. $K(p_i, t)=\{k_1, k_2, \ldots, k_m\}$ where $t>=t(p_i)$. To simplify the problem, we ignore the user searching behavior which may bring additional influence and consider only $t\in\{t_1, t_2, \ldots, t_n\}$. In contrast to traditional keyword extraction, the problem addressed here is dynamic keyword extraction as $K$ apparently is a function of time $t$.

## 4. DYNAMIC KEYWORD EXTRACTION

In this section, we mainly focus on dynamic inter-post keyword extraction for inner ads. Basically, the task involves two logical steps, i.e. (1) when a new post $p_i$ is uploaded at time $t_i$, generate the keywords for the current new post $p_i$; (2) later on at time $t_j$ ($j>i$), update the keywords for the existing post $p_i$ when it is associated directly or indirectly to a new uploaded post $p_j$ through the post replying links. Before explaining the keyword generation and update processes, we first present the following assumptions and our word weighting strategies based on linking relations.

ASSUMPTION 1: A post $p_i$ contains the content in its main body $b_i$ and an optional subject line $s_i$. The importance of these two kinds of content should be differentiated. It is reasonable to assume that the information conveyed in the subject, if provided, is more important than the information conveyed in the main body since the subject can be considered as a human-written abstract of his/her post.

We define the weight of a word $w$ in a post $p_i$ at any time $t_k$ as

$$weight(w, p_i, t_k) = \alpha * weight(w, s_i, t_k) + weight(w, b_i, t_k) \quad (1)$$

where $\alpha>1$. The tf*idf weighting method is applied to estimate the weight of the word. When $p_i$ is first time uploaded at time $t_i$, the weights of the word $w$ in the subject $s_i$ and the main body $b_i$ are initiated as

$$weight(w, s_i, t_i) = tf{*}idf(w, s_i, t_i) \quad (2)$$

and

$$weight(w, b_i, t_i) = tf{*}idf(w, b_i, t_i) \quad (3)$$

$weight(w, b_i)$ and $weight(w, p_i)$ in turn may be revised later when another post, say $p_j$, replied to $p_i$ at time $t_j$ ($j>i$).

ASSUMPTION 2: Since a post $p_i$ is associated with an optional reply session $r_i$, we consider to extract the keywords from the extended scope of the post, i.e. in $p_i \cup r_i$. In other words, the keywords to be included in $K(p_i, t_k)$ are not necessarily limited to the words in $p_i$. However, the word $w$ in $p_i$ is normally considered more important than it is in $r_i$.

Based on this assumption, we calculate the weight of the word $w \in p_i \cup r_i$ at any time $t_k$ as

$$\begin{aligned} &weight(w, p_i, r_i, t_k) \\ &= weight(w, p_i, t_k) + \beta * weight(w, r_i, t_k) \end{aligned} \tag{4}$$

where $\beta$ <1 is a discount parameter. It means that the weight of the word is discounted when it appears in the reply session. In case of a NULL reply, $weight(w, r_i, t_k)$=0.

ASSUMPTION 3: When a post $p_i$ replies to an anterior post $p_j$ at time $t_i$, the two posts are usually closely related to each other. The content of $p_j$ may entirely or partially appear in the quotation $q_i$ of a reply session $r_i$ associated to $p_i$. Or it may be implied in $p_i$ if the quotation is not explicitly included. In the latter case, we assume that the post $p_i$ actually refers back to the whole content in $p_j$ implicitly.

Then, we calculate the weight of $w$ in the reply $r_i$ to the post $p_j$ at any time $t_k$, i.e. $weight(w, r_i, t_k)$ where $pr_i = p_j$, as follows.

If $q_i \neq$ NULL,

$$weight(w, r_i, t_k) = weight(w, p_j, t_k) \tag{5}$$

Otherwise,

$$weight(w, r_i, t_k) = \gamma * weight(w, p_j, t_k) \tag{6}$$

where $\gamma$ <1 is another discount parameter which differentiates implicit quotations from explicit quotations.

ASSUMPTION 4: The relation established by post reply is bi-directional. That is to say, when the post $p_j$ replies to the post $p_i$, not only does $p_j$ influence the weights of the words in the quotation of $p_j$, the weight of the words in $p_i$ is also influenced by the quotation in $p_j$.

This happens when the new post $p_j$ is uploaded and it replies to an existing post $p_i$. Then the weights of the words in $p_i$ are revised, actually, increased due to $p_j$'s interest and concern to $p_i$. Similar to Formulas (5) and (6), if $q_j \neq$ NULL, the weights of the words in $p_i$ that also appear in the quotation $q_j$ are revised as,

$$\begin{aligned} &weight(w, p_i, t_k) \\ &= weight(w, p_i, t_{k-1}) + \beta * weight(w, q_j, t_i) \end{aligned} \tag{7}$$

where $\beta$ <1. Notice that the increased weights are determined according to the original tf*idf weights of those words at $t_i$. Otherwise, for all the words in $p_i$,

$$\begin{aligned} &weight(w, p_i, t_k) \\ &= weight(w, p_i, t_{k-1}) + \beta * \gamma * weight(w, b_i, t_i) \end{aligned} \tag{8}$$

Figure 2 illustrates the processes of inter-post keyword generation and update and Algorithm 1 explains how to generate a set of keywords for a new post and when to update existing keyword set(s).

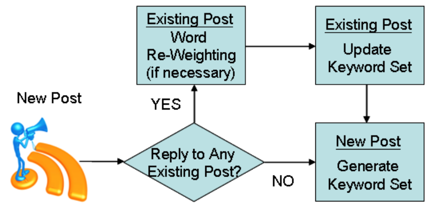


**Figure 2. Flowchart of Inter-Post Keyword Generation and Update Processes**

**Algorithm 1: Dynamic Inter-Post Keyword Extraction**

When a new post $p_i$ is uploaded at $t_i$,

1. IF $p_i$ replies to $p_j$ (i.e. $j<i$)
    1.1. Do the necessary **Word Re-Weighting**
        1.1.1. Revise the weights of the words in $p_j$, i.e. $weight(w, p_j, t_i)$, by Formula (7) or (8);
        1.1.2. For any $p_k$ that replies $p_j$, (i.e. $j<k<i$), revise the weights of the words in $r_k$, i.e. $weight(w, r_k, t_i)$, by Formula (5) or (6);
    1.2. **Update** $K(p_j, t_i)$ and $K(p_k, t_i)$
        1.2.1 Rank the words in $p_j$, and $p_k$ according to the weights $weight(w, p_j, r_j, t_i)$ and $weight(w, p_k, r_k, t_i)$ calculated by Formula (1) to (4);
        1.2.2. Select the top ranked $m$ terms as keywords in $K(p_j, t_i)$ and $K(p_k, t_i)$ if their weights are greater than a pre-specified threshold.
   ENDIF
2. **Generate** $K(p_i, t_i)$ by the steps similar to 1.2.1.-1.2.2.

In the remainder of this section, we address global keyword extraction for banner/sidebar ads. This time, the traditional approach is applicable. It may work like this. At any time $t$, the whole set of the posts that are available at $t$ are taken into account and the top ranked $m$ words based on the tf-idf weighting method are extracted. Notice that the keywords generated at different time are not necessarily the same.

Our dynamic keyword extraction approach is different from the traditional approach when dealing with global keyword extraction. Similar to dynamic inter-post keyword extraction, dynamic global keyword extraction also take the relations built upon replies and quotations into account. Differently, these relations are considered as unidirectional and are calculated only once for the weights of the words in the posts, in particular $\beta$=0 in Formula (4).

ASSUMPTION 5: The weight of a word in the whole page is the accumulated weights of it in all the posts on the page.

So, the weight of a word $w$ on a page $P$ at any time $t_k$ is calculated as

$$weight(w, P, t_k) = \sum_i weight(w, p_i, t_{k-}) \ (i=1, 2, ..., n) \tag{9}$$

In the case where the number of the qualified inter-post keywords is less than $m$, the top-ranked global keywords are added to $K(p_j, t_i)$ or $K(p_k, t_i)$ as a supplement.

## 5. EVALUATION AND DISCUSSION

The data set we use to evaluate the proposed approach is the topic pages linked from the topic list pages on a popular Chinese mobile phone forum called Dospy (http://bbs.dospy.com), which is a good example of the forums using Discuz solution[1]. Among 1841 pages crawled from Dospy, 892 pages (i.e. 48.45%) contain reply and quotation sessions. On 720 pages (i.e. 39.11%), the first reply posts are found in the top 30 posts, and 91 out of 720 pages (12.64%) have more than 5 replies in the top 30 posts.

We implement a dynamic advertising keyword extraction system in python. The pymmseg-cpp[2] is embedded in the pre-processor to segment Chinese sentences into word tokens. The word idf information is obtained from the SOGOU dictionary[3], which is collected from more than 100 million Chinese Web pages. The parameters in our preliminary experiments are set intuitively. The title weight parameter $\alpha$ is set to 2. The quotation weight discount parameters $\beta$ and $\gamma$ are both set to 0.5. We extract 6 global keywords for the whole page (3 for banner ads and 3 for sidebar ads) and 3 inter-post keywords for each individual post.

We evaluate the significance of keyword extraction in terms of the average keyword effect measured for all the keywords extracted for a page or the posts over time. For each keyword, an expert is invited to judge whether it is closely related to the topics mentioned in the posts around it. It is given a positive score 1 if the answer is yes. A negative score -1 is assigned to the keyword if it clearly brings a negative effect, such as the keyword that has no relevance to the posts at all and that may made users take displeasure against it. For those keywords not clearly relevant to the posts while showing no negative effects, the 0 scores are assigned.

Recall that the traditional extraction approach generates global keywords for the whole pages only. In order to evaluate the keywords extracted for individual posts dynamically, we simply choose the top 3 global keywords as the inter-post keywords for all the places on the page in the traditional approach. Considering the linking information plays an important role in the proposed dynamic extraction approach, all those 91 pages that have more than 5 replies are evaluated in order to examine whether the proposed approach is effective or not. For reference, we randomly select 25 topic pages from the pages without reply. The evaluation results are presented in Table 1.

**Table 1: Evaluation by keyword effect**

| Keyword Effect | Traditional | Dynamic |
|---|---|---|
| Global Keyword (on Pages without Reply) | 1.295 | 1.327 (+2.47%) |
| Inter-post Keyword (on Pages without Reply) | 0.874 | 0.906 (+3.66%) |
| Global Keyword (on Pages with Many Replies) | 1.654 | 1.776 (+7.38%) |
| Inter-post Keyword (on Pages with Many Replies) | 1.056 | 1.123 (+6.34%) |

It shows in Table 1 that the global and the inter-post keyword effects can both be improved by the dynamic approach. We believe that the use of linking information contributes much to the improvement. In addition, for inter-post keywords, the effect improvement of the dynamic approach over the traditional approach on the pages with many replies is more significant than that on the pages without reply. This further validates our assumption that the reply and quotation information are important for ads keyword extraction.

Here are some observations in the evaluation:

(1) It is quite common that the posts in the first half of a page talks about a mobile phone, and then the rest discusses which earphone is suitable. While the traditional approach cannot accommodate to this topic shift problem well, the dynamic approach can capture this change easily and thus can provide more relevant keywords.

(2) When a post mentions an interesting function about a phone, people may reply with a very short note like "I like it" or simply quote "Smile Shutter" to express his/her endearment to the function mentioned. When a post has less or even no words in its own post body, the traditional approach does not work well. The dynamic approach that makes use of the words in its quotation or in the post it replies to extract the keywords relevant to the post can do better.

(3) Sometimes, when a person talks about a couple of favorite functions of the camera on his mobile phone and gives a video clip to show how these functions work, the replies may be only interested in the video clip. The information implicit in post replies help to pick up the most important or to say the most concerned camera functions on a post by readers.

## 6. CONCLUSION

In this paper, we propose a dynamic ads keyword extraction approach to make use of the linking relations built upon replies and quotations. When evaluated in terms of keyword effect, the all-round improvement of the proposed approach over the transitional approach is observed. Especially, the effectiveness is more significant when the topics on a page shift more often or when the posts themselves contain less content words but include quotations and replies to the previous posts.

## 7. ACKNOWLEDGMENTS

The work presented in this paper was supported by a grant from Hong Kong RGC (No. PolyU5217/07E).

## 8. REFERENCES

[1] W. Yih, J. Goodman, and V. R. Carvalho. Finding advertising keywords on the web pages. In *WWW'06: Proc. of the 15th intl. conf. on World Wide Web*, pages 213-222, New York, NY, 2006.

[2] A. Broder, M. Fontoura, V. Josifovski, and L. Riedel. A Semantic Approach to Contextual Advertising. In *SIGIR'07*, pages 559-567, Amsterdam, The Netherlands, 2007.

[3] G. Mishne, and N. Glance. Leave a reply: An analysis of weblog comments. In *Proc. of WWW'06 Workshop on the Weblogging Ecosystem*, Edinburgh, UK, 2006.

[4] Meishan Hu, Aixin Sun, and Ee-Peng Lim. Comments-Oriented Blog Summarization by Sentence Extraction. In *CIKM'07*, pages 901-904, Lisboa, Portugal, 2007

.

[1] Notice that Dospy has already placed the contextual ads after the first post and the third post on each topic page.

[2] http://code.google.com/p/pymmseg-cpp/

[3] http://www.sogou.com/labs/dl/w.html