

# Finding Readings for Scientists from Social Websites

Jiepu Jiang, Zhen Yue, Shuguang Han, Daqing He

School of Information Sciences, University of Pittsburgh

jiepu.jiang@gmail.com, zhy18@pitt.edu, shh69@pitt.edu, dah44@pitt.edu

## ABSTRACT

Current search systems are designed to find relevant articles, especially topically relevant ones, but the notion of relevance largely depends on search tasks. We study the specific task that scientists are searching for worth-reading articles beneficial for their research. Our study finds: users' perception of relevance and preference of reading are only moderately correlated; current systems can effectively find readings that are highly relevant to the topic, but 36% of the worth-reading articles are only marginally relevant or even non-relevant. Our system can effectively find those worth-reading but marginally relevant or non-relevant articles by taking advantages of scientists' recommendations in social websites.

## Categories and Subject Descriptors

H.2.8 [Database Applications]: scientific databases.

## General Terms

Algorithms, Measurement, Performance, Experimentation.

## Keywords

Scientific articles; scientific readings; social search.

## 1. INTRODUCTION

Scientists in nowadays not only read articles closely related to their main research fields but also those from other fields. This makes automatically finding useful readings more difficult than simply matching articles with queries because: 1) many useful readings could come from topics that the scientists do not have adequate knowledge to formulate effective queries; 2) not all topically relevant articles are worth reading; and 3) some useful readings may not contain query terms.

We believe that a useful search system for solving this problem should have the following key features: 1) the scientists are not required to formulate queries related to the topics of the readings, but rather queries about the problems or research topics they are working on, which should be a relatively easier task; 2) the system can find not only readings that are topically relevant, but also those beyond the topics that the scientists are working on. Here we propose a method of finding readings by looking at what scientists' peers are reading about in social websites, which generates a list of candidate readings not limited to topically relevant articles. Specifically, we study the following research questions:

(1) What are the relations between users' perception of relevance and preference of reading?

(2) Can peer scientists' libraries help find worth-reading articles beyond the queries' topics?

## 2. EXPERIMENT DESIGN

### 2.1 Dataset

We built a dataset [1] based on articles and users' libraries in CiteULike, a social reference management website. In CiteULike,

each user can maintain a collection of articles as the user's personal library. Here we assume CiteULike users are scientists and their personal libraries are their collections of useful readings. Previous studies found co-occurrences of articles in users' libraries can be used for clustering articles into research fields [2]. Our dataset includes titles and abstracts for 913,846 unique articles posted to CiteULike by users in 2010 (99.2% of all articles posted by users in 2010) and 54,402 users' personal libraries.

### 2.2 Algorithms

Let  $R$  be an article (reading). Given a query  $q$ , we rank readings by  $P(R|q)$ , which is equivalent to  $P(R, q)$  in ranking. Further, we model  $P(R, q)$  as: what is the probability that a scientist working on the problems described by  $q$  will read a reading  $R$ ?  $P(R, q)$  is calculated by two steps. First, we find a list of peer scientists by  $P(q|u)$  and  $P(u)$ :  $P(u)$  is considered equal for all users, i.e.  $1/|u|$ ;  $P(q|u)$  is calculated as (2) using expert finding "model 2" in [3]. In (2),  $L_u$  is the list of articles in  $u$ 's personal library, which is used to model  $u$ 's expertise. Second, we let each peer scientist  $u$  vote for reading  $R$  by the probability  $P(R|u, q)$ . We assume a peer scientist  $u$  will vote for  $R$  as a worth-reading article for  $q$  if  $R$  is in both  $L_u$  and the search results of query  $q$  (referred to as  $S_q$ ), as in (3). The whole method is referred to as RUL, which is summarized in (1). We can further set a cutoff value  $n$  in (3) so that only the top  $n$  retrieved results for query  $q$  are considered for voting (referred to as  $RUL_n$ ). We use two ad hoc search models for comparison: sequential dependence model (SDM) and relevance model 3 (RM3 [4]). We also merge rankings of  $RUL_n$  and  $SDM/RM3$  by (4).

$$\text{RUL: } P(R, q) = \sum_u P(R, q|u)P(u) = \frac{1}{|u|} \sum_u P(q|u)P(R|u, q) \quad (1)$$

$$\text{Finding peer scientists: } P(q|u) = \sum_{d \in L_u} P(q|d, u) \cdot P(d|u) \quad (2)$$

$$\text{Voting for articles: } P(R|u, q) = \begin{cases} 1 & \text{if } R \in L_u \text{ and } R \in S_q \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

$$\text{Merge rankings: } \text{Score}(R) = \frac{1}{\text{rank}(RUL_n)} + \frac{1}{\text{rank}(SDM/RM3)} \quad (4)$$

### 2.3 User Judgments

We recruited 10 subjects in academia for judgments (1 faculty member, 1 postdoc and 8 PhD students). Every participant was asked to generate 3 queries, each of which described a research problem related to his/her research. Seven runs were pooled (depth=10) for each query: SDM, RM3, RUL, and  $RUL_n$  with  $n = 20, 50, 100, \text{ and } 200$ . On average 31 articles were pooled for each query. The participants need to answer the following questions:

Q1. Do you think this article is relevant to your query? (1-not relevant, 2-somewhat relevant, 3-relevant)

Q2. Do you want to read this article? (1-I already read it, 2-yes, I want to read it, 3-no, I don't want to read it)

Q3. (If Q2=1) I already read it (1-I like it, 2-neutral, 3-I don't like it)

For each query, after judging all pooled articles, we asked the participant to select at most 3 articles he/she wanted to read first if he/she is only given limited time. We assign relevance scores ( $rel$ ) to articles based on answers to Q1 ( $rel = 2$  for "relevant";  $rel = 1$  for "somewhat relevant"; and  $rel = 0$  for "not relevant"). For read-

ing preference score (*read*), we consider two cases: if the participant has read the article ( $Q2=1$ ), we assign *read* score based on  $Q3$  ( $read = 2$  for “I like it”;  $read = 1$  for “neutral”; and  $read = 0$  for “I don’t like it”); if the participant has not read the article, we assign  $read = 2$  to the three articles the participant selected to read first, and  $read = 1$  to other articles the participant would like to read ( $Q2=2$ ). Table 1 shows statistics of participants’ judgments. For the 477 articles judged as worth-reading ( $read \geq 1$ ), 173 (36%) are only marginally relevant or not relevant to the queries’ topics ( $rel \leq 1$ ). After removing 370 articles that are neither relevant to the topic nor worth-reading ( $rel = 0; read = 0$ ), we find a moderate correlation ( $r = 0.357$ ) between articles’ *rel* and *read* scores. This indicates the fact that good readings are not necessarily relevant to the queries’ topics and vice versa.

**Table 1. Article judgments statistics.**

	rel = 2	rel = 1	rel = 0	sum
read = 2	166	40	6	212
read = 1	138	111	16	265
read = 0	13	76	370	459
sum	317	227	392	936

### 3. RESULTS AND DISCUSSIONS

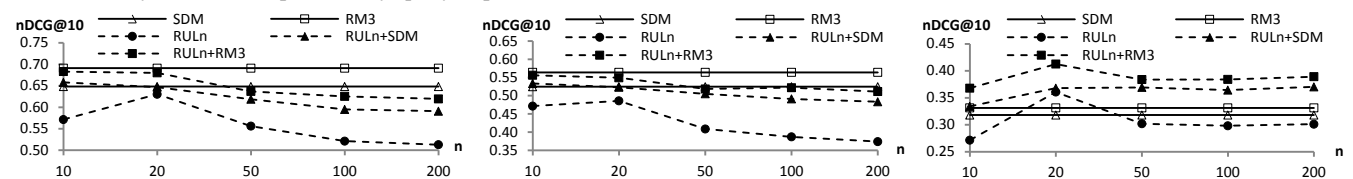
Our evaluation includes two stages:

(1) At the first stage, we use readings that are highly relevant ( $rel = 2; read \geq 1$ ) for evaluation (**HRelRead**). As shown in Table 1, over 90% of the highly relevant articles ( $rel = 2$ ) are also worth-reading ( $read \geq 1$ ). Current search systems (SDM and RM3) may perform well in finding HRelRead articles because the HRelRead readings are mostly also highly relevant articles. We also evaluate by article relevance (**QRel**) for comparison.

Left and center parts of Figure 1 show the results using QRel and HRelRead for evaluation. Ad hoc search models (SDM and RM3) performed very effectively ( $nDCG@10 \geq 0.6$  and  $0.5$ ). In comparison,  $RUL_n$  performed worse than ad hoc search models, but is still very effective (when  $n = 20$ ,  $nDCG@10 \geq 0.45$ ). This is not surprising considering *read* and *rel* scores are highly correlated in HRelRead readings.  $RUL_n$  does not explicitly model topical relevance, but SDM and RM3, as two of the best ad hoc search models, model topical relevance very well. Thus, in general, ad hoc search models can effectively find the HRelRead readings, but their success may come from better modeling of topical relevance rather than better modeling the task of finding readings.

(2) At the second stage, we use readings that are marginally relevant or not relevant for evaluation (referred to as **MRelRead**;  $rel \leq 1, read \geq 1$ ). Since such readings are not necessarily relevant, ad hoc search models may not perform effectively. The right part of Figure 1 shows evaluation results for MRelRead readings. It indicates that ad hoc search algorithms (SDM and RM3) cannot perform as effectively as they did in finding HRelRead readings ( $nDCG@10 < 0.35$ ). In comparison, when  $n = 20$ ,  $RUL_n$  can perform significantly better than SDM ( $nDCG@10+13.36\%; p < 0.05$ ). Combining  $RUL_n$  with SDM/RM3 further improved performance.

As expected, ad hoc search models cannot effectively solve the problems of finding MRelRead readings because in such cases article relevance and user’s preference of reading diverge a lot. It seems unlikely to solve the problem by query expansion, as RM3

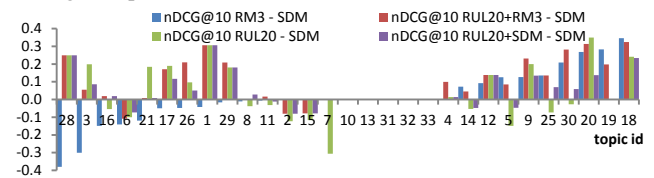


**Figure 1. nDCG@10 of SDM, RM3, and  $RUL_n$  (+SDM/RM3) evaluated by QRel (left), HRelRead (center), and MRelRead (right).**

performed not much better than SDM. The reasons can be: 1) if the original query is not effective, it is unclear whether or not pseudo-relevance feedback can produce high quality expanded terms; and 2) expanded terms may enrich representations of the query’s topic, but do not necessarily help matching of cross-disciplinary articles.

Figure 2 shows per topic differences of  $nDCG@10$  between SDM, RM3,  $RUL_{20}$ , and  $RUL_{20}+SDM/RM3$ . It shows that RM3 improves and hurts nearly the same number of topics compared with SDM, but  $RUL_{20}$  (+SDM/RM3) can not only improve topics hurt by RM3, but can also get as much as improvements in topics improved by RM3. We find the reasons are related to the voting mechanisms in  $RUL_n$ . On the one hand, voting can reduce wrong expansions of topics. For example, for topic 28 “science mapping intellectual structure”, SDM performs effective but RM3 wrongly emphasizes on “mapping” and matches articles such as “Genetic mapping in human disease”. However, since peers found by  $RUL_n$  do not read and vote for the article, it is still lower ranked in results of  $RUL_n$ . On the other hand, voting is also a useful way of finding good readings beyond topical level. Still, for topic 28, the user wants to read the article titled “Scholarly research and information practices: a domain analytic approach”, which is not ranked in top position by SDM or RM3 for the lack of query terms in title and abstract. Instead,  $RUL_n$  votes a high score for this article because it has been read by many scientists. We also find, however,  $RUL_n$  works poorly if the topic is associated with peers with diversified interests, e.g. topic 7 “collaborative information seeking”.

To conclude, our evaluation results demonstrate that current search models can effectively find readings that are also topically relevant, because in such cases article relevance and preference of reading are highly correlated. In comparison, our system can more effectively find readings beyond the query’s topic by taking advantages of peer scientists’ recommendations in social websites.



**Figure 2. Per topic nDCG@10 differences of algorithms.**

### 4. ACKNOWLEDGMENTS

This work was supported in parts by National Science Foundation grant IIS-1052773 and III-COR 0704628.

### 5. REFERENCES

- [1] <http://www.sis.pitt.edu/~jjiang/data/rul/>
- [2] J. Jiang, D. He, C. Ni. 2011. Social reference: aggregating online usage of scientific literature in CiteULike for clustering academic resources. In JCDL '11: 401-402.
- [3] K. Balog, L. Azzopardi, M. de Rijke. 2006. Formal models for expert finding in enterprise corpora. In SIGIR '06: 43-50.
- [4] Y. Lv, C. Zhai. 2009. A comparative study of methods for estimating query language models with pseudo feedback. In CIMK '09: 1895-1898.