Fresh and Diverse Social Signals: Any Impacts on Search?

Ismail Badache, Mohand Boughanem

Fresh and Diverse Social Signals: Any Impacts on Search?

Ismail Badache
LSIS UMR 7296 CNRS
University of Aix-Marseille
Marseille, France
Ismail.Badache@lsis.org

Mohand Boughanem
IRIT UMR 5505 CNRS
University of Toulouse III
Toulouse, France
Mohand.Boughanem@irit.fr

ABSTRACT

In this paper, we extensively study the impact of social signals (users' actions) obtained from several social networks on search ranking task. Social signals associated with web resources (documents) can be considered as an additional information that can play a vital role to estimate a priori importance of these resources. Particularly, we are interested in the freshness of signals and their diversity. We hypothesize that the moment (the date) when the user actions occur and the diversity of actions may impact the search performance. We propose to model these heterogeneous social features as document prior. We evaluate the effectiveness of our approach by carrying out extensive experiments on two different INEX datasets, namely SBS and IMDb, enriched with several social signals collected from social networks. Our experimental results consistently demonstrate the interest of integrating fresh and diverse signals in the retrieval process.

Keywords

Social signals, Social networks, Signals diversity, Freshness.

1. INTRODUCTION

The majority of information retrieval (IR) systems exploit two classes of features to rank documents in response to user's query. The first class, the most used one, is query-dependent, which includes features corresponding to particular statistics of query terms such as term frequency, and term distribution within a document or in the collection of documents. The second class, referred to as documents prior, corresponds to query-independent features such as the number of incoming links to a document [26], PageRank [9], topical locality [15], presence of URL [35], document authors [27] and social signals [25].

DOI: http://dx.doi.org/10.1145/3020165.3020177

Among these features, social signals such as (*like*, *+1*, *share*, *tweet*, *comment*) are probably one of the most interesting sources of evidence to measure document prior. Indeed, different web pages use social network buttons which allow users to express their support by (*like*, *+1*), *recommend* content [1], *comment* or *rate* a resource, send a *tweet* mentioning a resource, etc. In 2016, some statistics show that among 3.7 billion Internet users 76% are registered on at least one social network[1]. Such significant statistics have attracted researcher to study the degree of activity of those users. For instance, on Facebook, every 60 seconds there are about 50,000 posts with more than 2.3 million of *like* performed on different documents. All these social signals are valuable to improve document relevance ranking in conventional text search.

Most of existing approaches [11, 12, 22, 25] exploit these signals to estimate the document prior by simply considering the quantity (the number) of signals related to a resource, or the polarity for some specific signals. In this paper, we hypothesize that the time (age) of signals (when they occur), signals diversity on a resource, and the resource age may affect on the prior estimation. Therefore, we assume firstly, that fresh (recent) user actions may indicate some recent interests toward the resource. Secondly, the number of signals on a resource depends on the resource age. In general, an old resource may have much more signals than a recent one. Thirdly, we consider that signals diversity can be seen as a clue showing the degree of interest toward a resource, beyond a given social network or a community.

Note that some works have already investigated the temporality of social signals and their diversity associated to web resources in poster papers [6, 5]. They only described some preliminary results and they did not deeply evaluate and analyze the results. This paper extends significantly their work in the following additional aspects:

- exploiting additional signals namely, multi-valued signal such as *rating* (number of *rating* and *rating* value),
- evaluating the impact of freshness of signals,
- evaluating the signals normalization with resource age,
- evaluating jointly the impact of diversity and freshness of signals on IR process.
- conducting new and extensive experiments on a standard collection, namely INEX SBS,
- comparing the signals and identifying those that are important to improve IR.

[1] www.blogdumoderateur.com/chiffres-internet-2016-une-minute/

The research questions addressed in this paper are summarized as follows:

- What is the effect of the freshness of each action of the same type of signal (e.g. age of each action of *rating*) on IR system performance?
- What are the best temporal-dependent signals that can enhance a search?
- What is the impact of signals' diversity on IR process?

The remaining part is organized as follows. Section 2 reviews some related work. Section 3 describes our social approach. In section 4, we evaluate the effectiveness of our suggested approach and discuss the results. Finally, we conclude the paper with mentioning some future directions.

2. BACKGROUND AND RELATED WORK

In this section, we report related work that has leveraged social signals to measure a priori relevance of a resource. We distinguish those they do not take into account time when exploiting the social signals and those they do.

2.1 Time-Independent Signals Approaches

Some works focus on analyzing different statistical patterns of UGC (user generated content) such as YouTube videos [13], or on how to improve IR effectiveness by exploiting these UGC, particularly users' actions, with their underlying social network [7, 11, 12, 22, 23, 25].

Cheng et al. [13] presented a detailed investigation of characteristics of YouTube videos such as number of *views*, *comment*, number of *like* and number of *dislike*. This work does not exploit the potential of these studied features in retrieval process. In the last years, some works have concentrated on studying the richness and the possible use of these user-generated characteristics in search. *Chelaru et al.* [11, 12] studied the impact of social signals (*like*, *dislike*, *comment*, etc) on the effectiveness of search on YouTube. They showed that, although the basic criteria using the similarity of query with video title and annotations are effective for video search, social criteria are also useful and have improved the ranking of search results for 48% queries. They also evaluated the impact of social feedback on YouTube videos retrieval by using the state-of-the-art learning to rank approaches with a greedy feature selection algorithm.

Other studies are interested in exploiting social features to improve IR on the web and on the social networks. *Karweg et al.* [22] proposed an approach that combines topical score and social score based on two factors: (i) the engagement intensity, which specifies how strongly a user has interacted with a document using social services; for example, both *clicking* on a link and *recommending* a restaurant to a friend are interactions/engagements; (ii) the trust degree which determines how intense the relation between the user and other individuals registered on the same social network based on the *popularity* feature, using PageRank algorithm. They showed that the ranking improves as long as the searcher adds more friends, or his friends create more content over time. They found that social results are available for most queries and usually lead to more satisfying results. Similarly, *Khodaei and Shahabi* [25] proposed a ranking approach exploiting several social factors including the relationships between document owners and querying user, the importance of each user and user action (*playcount*: number of times a user listens to a track on lastfm) performed on web documents. They have conducted an extensive experiments on "lastfm" dataset. They showed a significant improvement of socio-textual ranking compared to the textual only and social only approaches.

Some prior studies tackled the *popularity* prediction of online content. We here briefly review some of them including social media signals to compute a "document prior" or, rather, a *popularity* for web content estimation (e.g. news article, tweet, etc). For example, *Borghol et al.* [7] analyzed differences among YouTube videos that have essentially the same content (clones) but different *popularity* using a multilinear regression model. This *popularity* is estimated based on *view* count, uploader's *followers* count, number of *comments*, *likes* and *favourite* events and average *rating*. *Bandari et al.* [3] and *Hong et al.* [19] exploited textual features extracted from messages (e.g., *hashtags* or *URLs*), the topic of the message, and user related features (e.g., number of *followers*) to predict the *popularity* of news and tweets. More specifically, *Hong et al.* [19] used *retweets* as a measure of *popularity* of *tweet* and then applied machine-learning techniques to predict how often new messages will be re-tweeted. In their work, different features have been used, including the content of messages, temporal information, metadata of messages and users, and the user's social graph. *Castillo et al.* [10] proposed a linear regression model to predict the total number of visits to a news article, using social media reactions. Unlike news, tweets, and videos, tips are associated with specific venues, and tend to be much less ephemeral as they remain associated with the venue (and thus visible to users) for a longer time. Thus, the *popularity* of a tip may be affected by features of the target venue as well.

Finally, there are other studies initiated by Microsoft Bing researchers [28, 33], which show the usefulness of different social contents generated by the network of user friends on Facebook. *Kazai and Milic-Frayling* [23] incorporated different types of social approval *votes* for book search using external resources that refer to books in the corpus, such as lists from libraries and publishers, and lists of bestsellers and award winning books. They defined a set of features to compute the social static rank and then train a neural network to integrate it with full-text search. They observed the effect of individual features with showing that the representations of the general consumer appeal tend to be more effective. Also, they found that social approval *votes* can improve a BM25F baseline that indexes both full-text and MARC[2] records. *Pantel et al.* [29] studied the leverage of social annotations on the quality of search results. They observed that the social annotations can benefit web search in two aspects: (i) the annotations are usually good summaries of corresponding web pages; (ii) the annotations indicate the *interest* and *popularity* of web pages. They also considered the type of social data (e.g., *like, share*) that can affect the user's choice. It was found that the user can benefit from such information in different ways such as a) personalized search results, b) participation in the activities of friends, and c) ranking results. These approaches exploit social signals of the experimental dataset, whereas our approach utilizes external signals from multiple social networks.

2.2 Time-Aware Social Signals Approaches

While considerable work has been done in the context of temporal query classification, there is still lack of studies that would analyze users' actions in temporal search from

[2]http://www.loc.gov/marc/

diverse viewpoints. Majority of existing works do not consider the time of action in the search process.

The works that are most related to our approach include [16, 21, 24], which attempt to improve ranking in web search. The approaches put forward by *Dong et al.* [16] and *Inagaki et al.* [21] used user click feedback features to identify how document relevance varies over time. More precisely, *Dong et al.* [16] incorporated fresh URLs extracted from Twitter into a general web search system. Using the labeled *<query,url>* training data pairs, a machine-learning ranking algorithm can predict the appropriate ranking of the search results for unseen queries. *Inagaki et al.* [21] proposed to exploit the temporal click, called ClickBuzz, which captures the interest of document over time. This method helps to exploit the user feedback to improve machine learning recency ranking by favoring URLs that have recent interest for the user's recency-sensitive query. The use of ClickBuzz in the ranking models leads mainly to an improvement in NDCG. *Khodaei and Alonso* [24] considered that the great mass of user-generated content in social networks provides an opportunity to examine how users produce and consume this type of content over time. They categorized the social interests of users into five classes: "recent", "ongoing", "seasonal", "past" and "random" and then analyze Twitter as well as Facebook data on social activities of users. They also discussed three different solutions where these time-sensitive signals can be applied: a) personalized IR; b) IR based on friends; and c) collective IR.

Another work presented by *Yuan et al.* [38] related to [19], and also heavily relies on temporal features to predict who will retweet a tweet. They investigated the dynamics of dyadic friend relationships through online social interactions, in terms of a variety of aspects, such as *reciprocity*, *temporality*, and *contextuality*. They propose a model to predict *repliers* and *retweeters* given a particular tweet posted at a certain time in a microblog-based social network. More specifically, they used learning-to-rank approach to train a ranker that considers user-level and tweet-level features (*like sentiment*, *self-disclosure*, and *responsiveness*) to address these dynamics. In the prediction phase, a tweet posted by a user is deemed a query and the predicted *repliers/retweeters* are retrieved using the learned ranker. *Borisov et al.* [8] was among the first to propose that time elapsed between a pair of user actions depends on the context of behaviors. They further construct a context-aware model to predict time between user actions in contexts. Their work shows that the dwell time of user clicks is affected by many different factors and incorporating such information may help the behavior model to better correlate with users' practical actions.

In this paper, we exploit novel social characteristics based on the same principle that some related work. However, unlike previous work [3, 7, 11, 12, 19, 23, 25], that attempt to improve a search on specific social networks (e.g. YouTube, Twitter) by exploiting their own signals, our work specifically focuses on, firstly, exploiting various signals from different sources as document priors to enhance Web IR. Secondly, considering *diversity* of signals as an additional factors in the estimation of the resource relevance. We note that in related work diversity has been applied only to the textual content of the document [2, 32]. Thirdly, evaluating the impact of the *freshness* of signals on the search performance by using their *creation date*. Fourthly, normalizing the distribution of signals on the resource using the *age of the resource*. Furthermore, we evaluated jointly the impact of *diversity* and *freshness* of signals on IR performance.

In addition, different to the approaches presented in [22, 29, 38], our approach does not take into account the user aspect, as well as we do not use a linear combination. However, we incorporate social signals and their different aspects (*freshness*, *diversity* and signal normalization with *resource age*) into a language model that provides a theoretical founded way to take into account the notion of a prior probability of a document. Finally our approach is completely unsupervised, and it is evaluated on different types of test data (Social Book Search and Internet Movie Database).

3. SOCIAL IR APPROACH

Our approach consists of exploiting social signals as a priori knowledge to be taken into account in retrieval model. We rely on language model to combine topical relevance of a given resource to a query and its importance modeled as a prior probability.

3.1 Notation

Social information that we exploit within the framework of our model can be represented by 5-tuple <U, R, A, T, SN> where *U, R, A, T, SN* are finite sets of instances: *Users*, *Resources*, *Actions*, *Times* and *Social Networks*.

Resources. We consider a collection $C=\{D_1, D_2, ...D_n\}$ of n documents. Each document (resource) D can be a web page, video or other type of web resources. We assume that resource D can be represented both by a set of textual keywords $D_w=\{w_1, w_2, ...w_z\}$ and a set of social actions A performed on this resource, $D_a=\{a_1, a_2, ...a_m\}$.

Actions. We consider a set $A=\{a_1, a_2, ...a_m\}$ of m actions (signals) that users can perform on resources. These actions represent the relation between users $U=\{u_1, u_2, ...u_h\}$ and resources C. For instance, on Facebook, users can perform the following actions: *like*, *share*.

Time. T represents two types of temporal dimensions:

1. The history of social action, let $T_{a_i}=\{t_{1,a_i}, t_{2,a_i}, ..t_{k,a_i}\}$ a set of k moments (date) at which action a_i was produced. A moment t_{k,a_i} represents the datetime of action a_i.
2. Age of resource, let $T_d=\{t_{D_1}, t_{D_2}, ..t_{D_n}\}$ a set of n dates at which each resource D was published. t_D is the publication date of the resource D, date is provided in datetime format.

3.2 Query Likelihood and Document Prior

We exploit language models (LM) [31] to estimate the relevance of document to a query. The language modeling approach computes the probability $P(D|Q)$ of a document D being generated by query Q as follows:

$$P(D|Q) \stackrel{\text{rank}}{=} P(D) \cdot P(Q|D) = P(D) \cdot \prod_{w_i \in Q} P(w_i|D) \quad (1)$$

$P(D)$ is a document prior, i.e. query-independent feature representing the probability of seeing the document. The document prior is useful for representing and incorporating other sources of evidence in the retrieval process. w_i represents words of query Q. Estimating $P(w_i|D)$ can be performed using different models (Jelineck Mercer, Dirichlet). The main contribution in this paper is how to estimate $P(D)$ by exploiting social signals.

3.3 Document Prior

Document prior $P(D)$ can be estimated by considering different assumptions. First, each action can be considered individually. In this case, we got as much probabilities as actions. Each $P(D)$ measures the impact of a given action relatively to the other actions in the document or in a set of documents. Second, we compute the joint effect of a set (group) of observed actions in a document, where each document is associated with one probability. In our case, we assume that the signals are independent. Therefore, the general formula is:

$$P(D) = \prod_{a_i \in A} P(a_i) \tag{2}$$

The set of actions A can be associated to a given group of signals. For instance, signals from a given social network (e.g. *TotalFacebook* which includes *like*, *share* and *comment*), or signals of the same type, or all signals. $P(a_i)$ estimation depends on the type of signal a_i: "mono-valued signals" such as *like*, *share* and *tweet* or "multi-valued signals" such as *rating*.

3.3.1 Mono-valued signals

In this case, the $P(a_i)$ is estimated using maximum-likelihood:

$$P(a_i) = \frac{Count(a_i, D)}{Count(a_\bullet, D)} \tag{3}$$

To avoid zero-valued probability, we smooth $P(a_i)$ by collection C using Dirichlet [39]. The formula becomes as follows:

$$P(a_i) = \frac{Count(a_i, D) + \mu \cdot P(a_i|C)}{Count(a_\bullet, D) + \mu} \tag{4}$$

$P(a_i|C)$ is estimated by using maximum-likelihood:

$$P(a_i|C) = \frac{Count(a_i, C)}{Count(a_\bullet, C)} \tag{5}$$

Where:

- $Count(a_i, Y)$ represents the number of actions a_i performed on Y (Y is either the document D or the Collection C).
- $Count(a_\bullet, Y)$ is the total number of actions performed on Y.

3.3.2 Multi-valued signals (rating)

Rating cannot be treated by a simple counting of actions as described above. *Rating* is a point in a range of values. For instance, values between 1 and 5 where 3 means "acceptable" and 5 "outstanding", and depends on the number of users who rate the document.

For this purpose, we use the Bayesian Average (BA) of the *ratings* [36] as a document prior, which takes into account the number of users that have rated the document and the rating value. Hereby, the BA of a document is computed as follows:

$$BA(D) = \frac{avg(r) \cdot |r| + \sum_{D' \in C} avg(r') \cdot |r'|}{|r| + \sum_{D' \in C} |r'|} \tag{6}$$

Where:

- $r = \{r_i\}$ is a set of *ratings* values associated to document D. r_i is the i^{th} *rating* given by user i to D.
- *avg* is the average value of *ratings* r associated to D.
- r' is the set of *ratings* in the collection C.

We note that considering logarithmic priors helps to compress the score range and thereby reduces the impact of the priors on the global score.

$$P(a_i|a_i = rating) = \frac{\log(1 + BA(D))}{\log(1 + \sum_{D' \in C} BA(D'))} \tag{7}$$

For documents with no *ratings*, this would result zero probability. In order to avoid this problem, we use the Add-One smoothing method:

$$P(a_i|a_i = rating) = \frac{1 + \log(1 + BA(D))}{1 + \log(1 + \sum_{D' \in C} BA(D'))} \tag{8}$$

Remark: we believe that this simple counting of signals may boost old resources compared to recent ones, because resources with long life on the web has much more chance to get more signals than recent ones. We need therefore to normalize the counts by the age of the resource. In addition, we assume that resources that have recent signals are more likely to interest the user. We propose to estimate the social importance of a resource by exploiting the moment when the interaction (signal) has occurred and the publication date of the resource. We take the same general model as described above. Also, we propose to take into account the temporal aspect, the time of the actions and the age of the resources. We describe in the following, how these two temporal aspects are considered.

3.4 Time-Based Document Prior

We assume that the resources associated with fresh signals (recent) should be favored (promoted) over those that are associated with old signals. Whenever a signal occurs, it is associated with its date of occurrence. We propose to associate to each action a temporal weight related to the time when it was associated to a resource. The corresponding formula is as follows:

$$Count_{t_a}(t_{j,a_i}, D) = \sum_{j=1}^{k} f(t_{j,a_i}, D) \tag{9}$$

The importance of signal at a given time can be estimated in different ways. A simple way is to take the inverse distance between actual date and the date of action, or an exponential weighting that further promotes the "recent" signals against the "old" signals:

$$f(t_{j,a_i}, D) = \exp\left(-\frac{\| t_{actual} - t_{j,a_i} \|^2}{2\sigma^2}\right) \tag{10}$$

We use formula 10 where $f(t_{j,a_i}, D)$ represents the signal-time function, estimated by using a Gaussian kernel [34]. This function calculates the temporal distance between the actual date t_{actual} and the date of the action t_{j,a_i}. $\sigma \in \mathbb{R}^+$ is the Gaussian kernel parameter.

The prior probability $P(D)$ is estimated using formula 4 but by replacing $Count()$ by $Count_{t_a}()$. We note that if the date of signal is ignored $f(t_{j,a_i}, D) = 1 \; \forall t_{j,a_i}$.

Regarding the *rating*, we consider the time when the vote occurred. We use the similar principle as above, the temporal *rating* is defined by using the Gaussian kernel as follows:

$$r_t = r_i \cdot \exp\left(-\frac{\| t_{actual} - t_{r_i} \|^2}{2\sigma^2}\right) \tag{11}$$

Where: $r_t = \{r_{t_i}\}$ is a set of temporal *ratings* associated to D. r_{t_i} is the i^{th} *rating* biased by it creation date t_{r_i} given by the user i to D.

We note that we estimate $P(a_i|a_i = rating)$ by using the formula 8, with replacing r by r_t in formula 6.

3.5 Signal Normalization with Resource Age

The priors we defined are proportional to the quantity of signals associated to a resource. This may lead to an undesirable effect where old resources will naturally have much more signals than new ones. This will promote old resources. To cope with this issue, we propose to normalize the distribution of social signals associated with a resource through resource publication date. Different functions can be used to estimate the normalization factor that takes into account the age of the resource. We used the Gaussian formula which will boost recent resource. The formula is the following:

$$A(D) = \exp\left(-\frac{\| t_{actual} - t_D \|^2}{2\sigma^2}\right) \quad (12)$$

The counting function $Count_{t_D}(a_i, D)$ and the $BA(D)$ will be normalized as follows:

$$Count_{t_D}(a_i, D) = Count(a_i, D) \cdot A(D) \quad (13)$$

$$BA_{T_D}(D) = BA(D) \cdot A(D) \quad (14)$$

The prior probability $P(a_i)$ based on normalization factor is estimated using formula 4, by replacing $Count()$ by $Count_{t_D}()$ for document and $Count_{t_C}()$ for collection, and for $P(a_i|a_i = rating)$ using the formula 8, with replacing $BA(D)$ by $BA_{T_D}(D)$.

3.6 Diversity of Social Signals

The diversity of signals associated with a resource reflects the variety of social communities which interact with this resource. It is important to mention that users are from heterogeneous social network and do not have the same social signals buttons. These diverse signals from different origins that judge the resource, can be play a role to cope the risk of manipulation and blind-following on the positive or negative judgment of content. For example, LinkedIn users do not see what Facebook users think about the same resource.

We believe that the diversity of signals on a resource is a clue that may indicate an interest toward a social network or a community. Diversity and quantitative distribution of social signals in a resource may be considered as factors of significance, i.e., a resource dominated by a single signal should be disadvantaged against a resource with an equitable distribution of the signals. Therefore, we propose to integrate signals diversity factor in the search model. We propose to evaluate this diversity by using diversity index of Shannon-Wiener [30], an entropy. It is given by the following formula:

$$Diversity(D) = -\sum_{i=1}^{m} P(a_i) \cdot log(P(a_i)) \quad (15)$$

Where $P(a_i)$ can be estimated as above (with or without considering *resource age*), and m represents the total number of signals.

The Shannon index is often accompanied by Pielou evenness index [30]:

$$Diversity^{Equit}(D) = \frac{Diversity(D)}{MAX(Diversity(D))} \quad (16)$$

Where:

$$MAX(Diversity(D)) = log(m) \quad (17)$$

The priori probability $P(D)$ is estimated using the formula 4 multiplied by the diversity factor as follows:

$$P(D) = \left(\prod_{a_i \in A} P(a_i)\right) \cdot Diversity^{Equit}(D) \quad (18)$$

4. EXPERIMENTAL EVALUATION

To evaluate our approach, we conducted a series of experiments on two datasets, SBS (Social Book Search) and IMDb (Internet Movie Database).We compared our approach which takes into account diversity and temporality of signals, with the baseline formed by only a textual model or simply counting of signals. Our goals in these experiments are:

1. to evaluate the impact of temporally-aware signals.
2. to evaluate the impact of signals diversity.

4.1 Description of Test Datasets

We used the collections SBS[3] and IMDb[4] documents provided by INEX. Each document describes a book on SBS and movie on IMDb. It is represented by a set of metadata, which has been indexed according to keywords extracted from fields (e.g. *title*, *actors* for IMDb and *book*, *summary* for SBS). For each document, we collected specific social signals via their corresponding API of 6 social networks listed in Table 3. We chose 30 and 208 topics with their relevance judgments provided by INEX IMDb 2011 and SBS 2015, respectively. In our study, we focused on the effectiveness of the top 1000 results.

Tables 1 and 2 show an example of social data of SBS and IMDb document. For instance, in Table 1, the IMDb document having Id=*tt1922777* was shared for the last time on Facebook at *2014-09-29T02:49:01* and published for the first time at *2011-05-07T19:00:57*. While, in Table 2 which contains SBS document, we have additional information compared to IMDb documents, the date of each *rating* associated with a document. For example, the SBS document having Id=*0553583859*, its first two *ratings* occurred on *1999-12-15* and *2000-01-04*.

Table 3 presents the social signals that we take into account in our experimentation.

Unfortunately, until now, Facebook does not allow the extraction of the dates of the various actions through its API. We have only the date of each *rating* from Amazon/LT.

Therefore, we represent the results using the formula 9 biased (taking account of the time) only by the date of *ratings*.

4.2 Results and Discussions

We conducted experiments with models based only on textual content of documents (Lucene Solr model, BM25 and Hiemstra language model without prior), as well as approaches combining textual content and social characteristics with taking into account their temporal aspect and diversity. We note that the best value of μ (smoothing parameter used in Dirichlet) belongs to the intervals: $\mu \in [90, 100]$ for IMDb and $\mu \in [2400, 2500]$ for SBS.

[3]http://social-book-search.humanities.uva.nl/#/data
[4]https://inex.mmci.uni-saarland.de/tracks/dc/2011/

Table 1: Instance of IMDb document with social data

	Facebook			Google+	Delicious	Twitter	LinkedIn
Id	Like	Share	Comment	+1	Bookmark	Tweet	Share
tt1922777	14763	13881	22914	341	12	2859	14
	Facebook						
Id	Last Share		Last Comment		Publication Date		
tt1922777	2014-09-29T02:49:01		2014-09-28T00:41:01		2011-05-07T19:00:57		

Table 2: Instance of SBS document with social data

	Facebook					
Id	Like	Share	Comment	Last Share	Last Comment	Publication Date
0553583859	137	60	17	2014-03-10T32:01:32	2014-03-18T00:01:43	2008-12-14T02:13:22
	Amazon/LibraryThing					
Id	Tag		Rating		Rating Date	
0553583859	13		4		1999-12-15	
			3		2000-01-04	

Table 3: Exploited social signals According Dataset

Social signals	Sources	Dataset
Number of *Comment*	Facebook	SBS, IMDb
Number of *Tweet*	Twitter	IMDb
Number of *Share(LIn)*	LinkedIn	IMDb
Number of *Share*	Facebook	SBS, IMDb
Number of *Reviews*	Amazon/LT	SBS
Number of *Tags*	Amazon/LT	SBS
Number of *Like*	Facebook	SBS, IMDb
Number of *Mention* +1	Google+	IMDb
Number of *Bookmark*	Delicious	IMDb
Number of *Ratings*	Amazon/LT	SBS

Remark: we have already shown taking into account these signals regardless of temporal aspect and diversity improves the IR performance, compared to baseline (based only on the topical relevance) [4] (Baseline (B) in Tables 4 and 5).

We can clearly notice in Tables 4 and 5 that considering signals separately or grouped (*All Criteria*) improve significantly the results. Furthermore, the obtained results using signals awarded first place in the INEX 2015 Social Book Search Competition[5]. This work is conducted in collaboration with *blinded* [20].

The different results are listed in tables 4, 5, 6 and 7 in terms of precision@k with $k \in \{10, 20\}$, nDCG and MAP. Baseline (A) represents textual models (BM25, Lucene Solr model and Hiemstra language model without prior), Baseline (B) represents configurations which do not take into account time in the priors estimation, and Baseline (C) represents configurations without considering signals diversity.

In order to check the significance of the results compared to the baseline, we conducted the Student's t-test [17] and the normality test [14]. We attached * (strong significance against Baseline (A), (B) and (C)) and ** (very strong significance against Baseline (A), (B) and (C)) to the performance number of each row in the tables when *p-value<0.05* and *p-value<0.01* confidence level, respectively. We discuss in the following the results of each configuration we investigated.

We evaluated our approach using different configurations, taking into account social signals as priors with: 1) their creation date, 2) publication date of the resource, and 3) the diversity of signals within a resource.

4.2.1 Impact of the "Time-Dependent Signal"

Table 4 lists the results obtained by integrating the date of the signal (the moment when a signal occurs). As aforementioned, the results obtained concern only the *rating* on the SBS. Table 4 show that the nDCG, precision@k and MAP are generally better compared to those obtained when the action time is ignored (Baseline (B) in Table 4). We record an improvement rate of 42.75% for *rating* compared to the run *BA_Rating* without considering time. These results are statistically significant.

[5] http://social-book-search.humanities.uva.nl/#/results15

Table 4: Results of P@k, nDCG and MAP for SBS

Models	P@10	P@20	nDCG	MAP
Baseline (A) : Without Priors				
BM25	0.0601	0.0517	0.1581	0.0517
Lucene Solr	0.0528	0.0487	0.1300	0.0463
ML.Hiemstra	0.0607	0.0559	0.1620	0.0527
Baseline (B) : Without Considering Time				
Like	0.0857	0.0689	0.1864	0.0741
Share	0.0901	0.0711	0.1900	0.0872
Comment	0.0799	0.0678	0.1807	0.0701
TotalFacebook	0.0958	0.0810	0.1937	0.0892
BA_Rating	0.0730	0.0559	0.1748	0.0620
Log_Tag	0.0770	0.0531	0.1742	0.0610
All Criteria	0.0973	0.0787	0.1981	0.0900
(a) With Considering Action Date T_a				
BA_Rating^{T_a}	0.0941**	0.0732**	0.1904**	0.0885**
(b) With Considering Publication Date T_D				
Like^{T_D}	0.0891*	0.0708*	0.1900*	0.0873*
Share^{T_D}	0.0917*	0.0796*	0.1947*	0.0903*
Comment^{T_D}	0.0881*	0.0711*	0.1882*	0.0777*
TotalFacebook^{T_D}	0.0957**	0.0873**	0.1959**	0.0928**
BA_Rating^{T_D}	0.0790*	0.0695*	0.1808*	0.0685*
Log_Tag^{T_D}	0.0782*	0.0599*	0.1771*	0.0666*
All Criteria^{T_D}	0.1078**	0.0973**	0.2080**	0.0986**

Table 5: Results of P@k, nDCG and MAP for IMDb

Models	P@10	P@20	nDCG	MAP
Baseline (A) : Without Priors				
BM25	0.3500	0.3371	0.4113	0.2068
Lucene Solr	0.3411	0.3122	0.3919	0.1782
ML.Hiemstra	0.3700	0.3403	0.4325	0.2402
Baseline (B) : Without Considering Time				
Like	0.3938	0.3620	0.5130	0.2832
Share	0.4061	0.3649	0.5262	0.2905
Comment	0.3857	0.3551	0.5121	0.2813
TotalFacebook	0.4209	0.4102	0.5681	0.3125
Tweet	0.3879	0.3512	0.4769	0.2735
+1	0.3826	0.3468	0.5017	0.2704
Bookmark	0.3730	0.3414	0.4621	0.2600
Share (LIn)	0.3739	0.3432	0.4566	0.2515
All Criteria	0.4408	0.4262	0.5974	0.3300
(b) With Considering Publication Date T_D				
Like^{T_D}	0.4091*	0.3620*	0.5308*	0.2907*
Share^{T_D}	0.4177*	0.3721*	0.5544*	0.2989*
Comment^{T_D}	0.3912*	0.3683*	0.5285*	0.2874*
TotalFacebook^{T_D}	0.4302	0.4258	0.5827	0.3200
Tweet^{T_D}	0.3918*	0.3579*	0.4903*	0.2779*
+1^{T_D}	0.3900	0.3511	0.5246	0.2748
Bookmark^{T_D}	0.3732	0.3427	0.4671	0.2618
Share^{T_D} (LIn)	0.3762	0.3449	0.4606	0.2542
All Criteria^{T_D}	0.4484*	0.4305*	0.6200*	0.3366*

We can explain these results as follows, for example, in the context of IMDb, searching for the name of an actor (e.g. topic number 2013106: *Paul Verhoeven Arnold Schwarzenegger*) the user might expect the most recent movies for the particular actor and social signals could help to discover them. However, we have not really evaluated the real impact of our proposal with respect to all the signals (except for the *rating*). But the results obtained by the $rating^{T_a}$ configuration, for which all conditions were satisfied are encouraging.

These results support our hypothesis, that states that the resources associated with fresh signals should be favored in comparison with those associated with old signals.

4.2.2 Impact of "Resource Age"

The results presented in part (b): *With Considering Publication Date* T_D of Tables 4 and 5, respectively, show that the nDCG and the precision are better compared to those obtained when the publication date of the resource is ignored (Baseline (A) and (B)). All signals either taken separately or grouped (according to their social networks (*TotalFacebook*), or all criteria) seem to take advantage from this normalization. These results seem to be correlated with the quantity of signals associated to the resource. For instance, the results for *share*, *like* and *comment* by considering the date are significantly better than those without considering date, in both collections. Indeed, we notice that the average number of *share* and *like* per resource are 362 and 360, respectively. Whereas the average number of *bookmark* is 13. The best results are obtained when combining all criteria (named $All\ Criteria^{T_D}$ in Tables 4 and 5), with a rate of improvement in terms of nDCG 4% on IMDb and 5% on SBS, compared to *All Criteria* where the time is ignored. Indeed, the resource that generates a social activity of 100 actions of *like* in one hour does not have the same importance and the same temporal interest for the users compared to a resource that gathered 100 of actions of *like* during a week.

4.2.3 Impact of "Diversity"

Tables 6 and 7 list the results when considering *TotalFacebook* and *All Criteria* with and without diversity. For both collections, the nDCG and the precisions are generally better where diversity is considered.

Table 6: Results of P@k, nDCG and MAP for IMDb

Models	P@10	P@20	nDCG	MAP
Baseline (C): Without Considering Diversity				
TotalFacebook	0.4209	0.4102	0.5681	0.3125
All Criteria	0.4408	0.4262	0.5974	0.3300
With Considering Diversity Div				
$TotalFacebook^{Div}$	0.4227*	0.4187*	0.5713*	0.3167*
$All\ Criteria^{Div}$	0.4463*	0.4318*	0.6174*	0.3325*

Table 7: Results of P@k, nDCG and MAP for SBS

Models	P@10	P@20	nDCG	MAP
Baseline (C): Without Considering Diversity				
TotalFacebook	0.0958	0.0810	0.1937	0.0892
All Criteria	0.0973	0.0787	0.1981	0.0900
With Considering Diversity Div				
$TotalFacebook^{Div}$	0.0960*	0.0840*	0.1945*	0.0907*
$All\ Criteria^{Div}$	0.1011*	0.0915*	0.2031*	0.0952*

The best results are obtained by combining all social criteria $All\ Criteria^{Div}$. This later, is better than both *TotalFacebook* and *All Criteria* without diversity factor. Taken into account all criteria (signals) coming from different social networks give more social credibility through the multitude of approval sources. Even though the quantity of signals coming from the other sources namely, LinkedIn, Google+, Delicious, are very low (in average 13, 29, 67 actions per resource, respectively) compared to those coming from Facebook (in average 360 per action and per resource), we notice that adding these sources improves significantly the results for both collections (P@10 is improved of 5% compared to $TotalFacebook^{Div}$ for both collections). Indeed, if many users of different social communities found that a resource is useful, then it is more likely that other users will find it useful too.

4.2.4 Impact of "Diversity" and "Resource Age"

Tables 8 and 9 summarize the results obtained when taking into account the diversity and signals normalization with a resource age. We list only the configurations that consider the grouped signals, the only ones for which diversity has meaning. We attached * (strong significance against the corresponding configurations which do not consider the publication date of the resource T_D and diversity Div).

Table 8: Results of P@k, nDCG and MAP for IMDb

Models	P@10	P@20	nDCG	MAP
Diversity and Document Publication Date T_D				
$TotalFacebook^{Div}_{T_D}$	0.4417*	0.4289*	0.5966*	0.3273*
$All\ Criteria^{Div}_{T_D}$	0.4568*	0.4334*	0.6311*	0.3427*

Table 9: Results of P@k, nDCG and MAP for SBS

Models	P@10	P@20	nDCG	MAP
Diversity and Document Publication Date T_D				
$TotalFacebook^{Div}_{T_D}$	0.0978*	0.0868*	0.1970*	0.0932*
$All\ Criteria^{Div}_{T_D}$	0.1092*	0.0988*	0.2095*	0.0997*

Comparing the results listed in Tables 8 and 9 with those obtained in the above (Tables 4, 5, 6 and 7), we notice that the diversity combined with the age of the resource factor improves the results on the two datasets. It seems that this combination takes cumulative advantage from both criteria, Diversity and Age of the resource, since this combination outperforms the results obtained by each criteria taken separately. Indeed, the resource that records a high diversity of signals during two days has not the same importance compared to a resource that records signals diversity during two weeks. In addition, a resource containing a high frequency of a variety of signals mean that the crowd judgment of this resource is more significant than a resource dominated by a single signal. A resource with a high signal diversity implies a high spread of this resource in several social networking communities.

We note that our method of combining diversity with the date of the action does not improve the results, therefore we have not presented the results of this configuration.

4.3 Diversity Distribution on Documents

In order to have a view on the distribution of diversity of social signals within documents, we have taken all documents returned by IMDb and SBS topics and then we have calculated the diversity score for each document. Then, we associated the relevance judgments provided by Qrels of INEX with each document. We represented the overlapping between the relevance and signals diversity through figures 1 and 2.

Figure 1: Signals diversity with respect to the relevance of IMDb documents

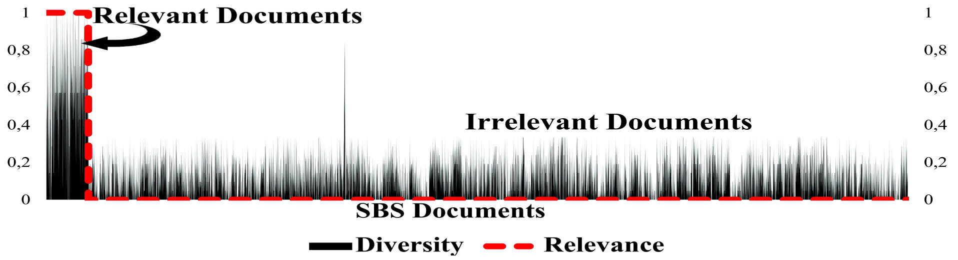

Figure 2: Signals diversity with respect to the relevance of SBS documents

Figures 1 and 2 show the scores of signals diversity for each document. In order to better differentiate between the relevant documents and the irrelevant ones, we sorted the documents according to their relevance. We note that the majority of relevant documents (returned) are delimited by the red rectangle in figures 1 and 2. They have the higher signals diversity scores compared to irrelevant documents, except one segment of irrelevant documents on SBS that shows high scores. We conclude that the diversity of signals obtains its best scores with the relevant documents.

4.4 Features Selection Algorithms Study

The problem of the described methods is in giving all signals same importance since we only count its frequency. However, it is obvious that signals are different. For instance a *like* signal designs a kind of positive vote for a given resource, whereas a *comment* might be positive or negative, in our case we only count this signal.

To better understand the importance of each signal, the real impact of social signals, we used the feature selection algorithms applied on both collections (IMDb and SBS). Feature selection algorithms [18] aim at identifying and eliminating as many irrelevant and redundant information as possible. Our goal is to determine the best time-dependent signals that can be effectively exploited in IR, as well as to verify if the results obtained previously (prior probability of the document) are consistent. We used Weka[6] for this experiment. It is a powerful open-source Java-based learning tool that brings together a large number of algorithms for selecting attributes.

We proceeded as follows, the top 1000 resources for each topic (30 IMDb topics and 208 SBS topics) were extracted using Lucene Solr model (using only text-based retrieval). Then, the scores (probability) of all criteria (signals) are calculated for each resource. We identify relevant resources and irrelevant according to the Qrels. The set obtained contains 30000 IMDb documents and 115248 SBS documents composed of:

- 2765 relevant documents and 27235 irrelevant documents for IMDb.
- 2953 relevant documents and 112295 irrelevant documents for SBS.

We observed that the classes of these sets are unbalanced. This occurs when there are many more instances in one class than others in a training collection. In this case, a classifier usually tends to predict samples from the majority class and completely ignore the minority class [37]. For this reason, we applied an approach to subsampling (reducing the number of samples that have the majority class) to generate a balanced collection composed of:

- 2765 relevant documents and 2765 irrelevant documents for IMDb.
- 2953 relevant documents and 2953 irrelevant documents for SBS.

Irrelevant documents for this study were selected randomly. Finally, we applied the attribute selection algorithms on the two sets obtained, for 5 iterations of cross-validation. **Remark**: these algorithms operate differently, some return an importance ranking of attributes, while others return the number of times that a given attribute has been selected by an algorithm in a cross-validation.

Tables 10 and 11 list the number of times (*folds number*: this value is between 0 and 5) that a given signal has been selected by an algorithm (9 signals were evaluated), and the rank of the signal (selected in order of preference) compared to other signals. Regarding the SBS, the value of *rank* is between 1 and 8, with 8 the number of evaluated criteria. For IMDb, the value of *rank* is between 1 and 9. For example, in Table 11 *tweet* has been selected 5 times by the algorithm *CfsSubsetEval* (numbered A_i=1), 3 times by *WrapperSubsetEval* and has been ranked 5^{th} among the 9 criteria by the algorithm *ChiSquaredAttributeEval*. We discuss the results in the following.

- **Impact of signals biased by age of document** : according to Tables 10 and 11 (columns corresponding to Publication date of document T_D), we note that

[6] http://www.cs.waikato.ac.nz/ml

Table 10: Selection of temporally dependent signals with attribute selection algorithms (Applied on SBS)

			Publication date of document T_D					Date of action T_a		
A_i	Algorithms	Metric	Comment	Share	Tag	Like	Rating	Comment	Share	Rating
1	CfsSubsetEval	[folds number]	5	5	2	5	2	3	5	5
2	WrapperSubsetEval	[folds number]	1	5	2	4	4	4	5	5
3	ConsistencySubsetEval	[folds number]	4	5	3	5	4	4	5	5
4	FilteredSubsetEval	[folds number]	4	5	2	5	2	3	5	5
		Average	3.5	5	2.25	4.75	3	3.5	5	5
5	ChiSquaredAttributeEval	[rank]	5	2	8	4	7	6	3	1
6	FilteredAttributeEval	[rank]	6	3	8	4	7	5	1	2
7	GainRatioAttributeEval	[rank]	6	1	8	4	7	5	3	2
8	InfoGainAttributeEval	[rank]	5	1	7	4	8	6	3	2
9	OneRAttributeEval	[rank]	6	3	7	4	8	5	2	1
10	ReliefFAttributeEval	[rank]	5	2	8	4	7	6	3	1
11	SVMAttributeEval	[rank]	5	2	8	4	7	6	3	1
12	SymetricalUncertEval	[rank]	5	2	8	4	7	6	3	1
		Average	5.375	2	7.75	4	7.25	5.625	2.625	1.375

Table 11: Selection of temporally dependent signals with attribute selection algorithms (applied on IMDb)

		Publication date of document T_D							Date of action T_a	
A_i	Metric	Comment	Tweet	Share(LIn)	Share	Like	+1	Bookmark	Share	Comment
1	[folds number]	5	5	0	5	5	3	0	5	5
2	[folds number]	3	3	2	5	5	3	2	5	2
3	[folds number]	5	5	5	5	5	5	4	5	5
4	[folds number]	5	5	0	5	5	3	0	5	5
	Average	4.5	4.5	1.75	5	5	3.25	1.5	5	4.25
5	[rank]	4	5	8	1	2	6	9	3	7
6	[rank]	5	4	8	1	2	6	9	3	7
7	[rank]	5	4	8	2	1	6	9	3	7
8	[rank]	4	5	8	2	1	6	9	3	7
9	[rank]	4	5	8	2	1	6	9	3	7
10	[rank]	5	4	8	1	2	6	9	3	7
11	[rank]	5	4	8	1	2	6	9	3	7
12	[rank]	5	4	8	1	2	6	9	3	7
	Average	4.625	4.375	8	1.375	1.625	6	9	3	7

the biased signals by the document age: *Share*(LIn), *Bookmark*, *Tag* and *Rating* are weakly favored by the selection algorithms with average ranks of 8, 9, 7.75, 7.25, respectively, and average selection of 1.75, 1.5, 2.25, 3, respectively. While in Table 11, the action +1 is moderately favored (with average ranks of 6 and average selection of 3.25). However, they are all selected by each algorithm except *Share*(LIn) and *Bookmark* in Table 11, which have not been selected by both algorithms *CfsSubsetEval* and *FilteredSubsetEval*. Indeed, this is an indication about their weak impact. The Facebook signals: *Like* and *Share* are the highest ranked (with averages ranks: 1.6 and 1.3, respectively) compared to other signals; they are strongly validated during the 5-fold cross-validation. The signals *Comment* and *Tweet* are in second place; they are often selected over the 5-fold cross-validation.

- **Impact of signals biased by their date** : regarding signals biased by their creation date, we remark through Table 10 that *Rating* is the best signal compared to all other signals. It is selected in 5 iterations of cross-validation by all algorithms and ranked first. One of the reasons of these results returns to the efficiency of exploitation of the dates of each action *Rating*. Consequently, our hypothesis is completely verified. We also note that *Share* comes in second position followed by *Comment* (with average ranks of 5.6 on SBS and 7 on IMDb). We recall that for these two signals, we have only used the date of the last action.

By comparing these results with the results appearing in Tables 4 and 5, we notice that the same factors of relevance highlighted by the model based on prior probability are highlighted by the study with attribute selection techniques. Social signals considered with their temporality, ($Rating^{T_a}$ and $Share^{T_D}$), which provide the best results (statistically significant) are the most favored and the highest ranked by the different selection algorithms. Finally, our study confirms the interest of the temporal dimension of social signals used to improve IR: relevant resources have a high number of social actions but also the recent interactions, and this number is proportional to the life of the resource on the web.

5. CONCLUSION

Previous work showed the impact of the signals without considering time. This paper focuses on studying the impact of *freshness* and *diversity* of signals associated with a resource on search. We have proposed to estimate the prior probability of document by considering these factors. The experiments conducted on IMDb and SBS datasets reveal that taking into account signals *diversity* and temporal aspects (*signal date, resource age*) in textual model improve the quality of search results. We used also feature selection algorithms to identify the best time-aware social signals for this task of IR.

For future work, we plan to address some limitations of the current study. We plan to study the importance of social networks and social actors of these signals and their impact on the relevance. This requires tracking users' personal profiles as well as those of their followers and those of users they *share*, *like, rate, tweet*, etc. We intend to collect these data in the future to evaluate the user preferences, compared to social neighbors, to solve the personalized search. User-centric data would likely help us for better understanding the temporal interests of users. Moreover, considering the polarity (positive, negative, neutral) of the textual content of the signals in search. This is even with these simple elements, the results encourage us to invest more this track.

6. REFERENCES

[1] O. Alonso and V. Kandylas. A study on placement of social buttons in web pages. *arXiv*, 2014.

[2] A. Angel and N. Koudas. Efficient diversity-aware search. In *SIGMOD*, pages 781–792, 2011.

[3] Roja B., Sitaram A., and Bernardo A.H. The pulse of news in social media: Forecasting popularity. In *ICWSM*, 2012.

[4] I. Badache and M. Boughanem. Social priors to estimate relevance of a resource. In *IIiX*, pages 106–114, 2014.

[5] I. Badache and M. Boughanem. Document priors based on time-sensitive social signals. In *European Conference on IR*, pages 617–622, 2015.

[6] I. Badache and M. Boughanem. A priori relevance based on quality and diversity of social signals. In *SIGIR*, pages 731–734, 2015.

[7] Y. Borghol, S. Ardon, N. Carlsson, D. Eager, and A. Mahanti. The untold story of the clones: content-agnostic factors that impact youtube video popularity. In *SIGKDD*, pages 1186–1194, 2012.

[8] A. Borisov, I. Markov, M. de Rijke, and P. Serdyukov. A context-aware time model for web search. In *SIGIR*.

[9] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *WWW*, pages 107–117, 1998.

[10] C. Castillo, M. El-Haddad, J. Pfeffer, and M. Stempeck. Characterizing the life cycle of online news stories using social media reactions. In *CSCW*, pages 211–223, 2014.

[11] S. Chelaru, C. Orellana-Rodriguez, and I. S. Altingovde. Can social features help learning to rank youtube videos? In *WISE*, pages 552–566, 2012.

[12] S. Chelaru, C. Orellana-Rodriguez, and I. S. Altingovde. How useful is social feedback for learning to rank youtube videos? *World Wide Web*, 2013.

[13] X. Cheng, C. Dale, and J. Liu. Statistics and social network of youtube videos. In *IWQoS Workshop*, pages 229–238, 2008.

[14] Ralph B. D'Agostino. Tests for the normal distribution. *Goodness-of-fit techniques*, 68:576, 1986.

[15] B.D. Davison. Topical locality in the web. In *SIGIR*, pages 272–279, 2000.

[16] A. Dong, R. Zhang, P. Kolari, J. Bai, F. Diaz, Y. Chang, Z. Zheng, and H. Zha. Time is of the essence: improving recency ranking using twitter data. In *WWW*, pages 331–340, 2010.

[17] William Sealy Gosset. The probable error of a mean. *Biometrika*, 6:1–25, 1908.

[18] Mark A. Hall and G. Holmes. Benchmarking attribute selection techniques for discrete class data mining. *IEEE Trans. on Knowl. and Data Eng.*, 15(6):1437–1447, 2003.

[19] L. Hong, O. Dan, and B. Davison. Predicting popular messages in twitter. In *WWW*, pages 57–58, 2011.

[20] M. Imhof, I. Badache, and M. Boughanem. Multimodal social book search. In *CLEF*, 2015.

[21] Y. Inagaki, N. Sadagopan, G. Dupret, A. Dong, C. Liao, Y. Chang, and Z. Zheng. Session based click features for recency ranking. In *AAAI Press*, 2010.

[22] B. Karweg, C. Huetter, and K. Böhm. Evolving social search based on bookmarks and status messages from social networks. In *CIKM*, pages 1825–1834, 2011.

[23] G. Kazai and N. Milic-Frayling. Effects of social approval votes on search performance. In *ITNG*, pages 1554–1559, 2009.

[24] A. Khodaei and O. Alonso. Temporally-aware signals for social search. In *SIGIR Workshop TAIA*, 2012.

[25] A. Khodaei and C. Shahabi. Social-textual search and ranking. In *CrowdSearch*, pages 3–8, 2012.

[26] W. Kraaij, T. Westerveld, and D. Hiemstra. The importance of prior probabilities for entry page search. In *SIGIR*, pages 27–34, 2002.

[27] C. Macdonald and I. Ounis. Voting for candidates: adapting data fusion techniques for an expert search task. In *CIKM*, pages 387–396, 2006.

[28] M. R. Morris and J. Teevan. Exploring the complementary roles of social networks and search engines. In *HCIC Workshop*, pages 1–10, 2012.

[29] P. Pantel, M. Gamon, O. Alonso, and K. Haas. Social annotations: Utility and prediction modeling. In *SIGIR*, pages 285–294, 2012.

[30] E.C. Pielou. Shannon's formula as a measure of specific diversity: its use and misuse. *American Naturalist*, pages 463–465, 1966.

[31] Jay M. Ponte and W. Bruce Croft. A language modeling approach to information retrieval. In *SIGIR*, pages 275–281, 1998.

[32] L. Qin, Jeffrey X. Y., and L. Chang. Diversifying top-k results. *VLDB Endowment*, 5(11):1124–1135, 2012.

[33] M. Raza. A new level of social search: Discovering the user's opinion before he make one. In *Microsoft Research*, pages 1–6, 2011.

[34] J-P Vert, K. Tsuda, and B. Schölkopf. A primer on kernel methods. *Kernel Methods in Computational Biology*, pages 35–70, 2004.

[35] T. Westerveld, W. Kraaij, and D. Hiemstra. Retrieving web pages using content, links, urls and anchors. 2002.

[36] Xiao Y. and Zhaoxin Z. Combining prestige and relevance ranking for personalized recommendation. In *CIKM*, pages 1877–1880, 2013.

[37] S-J. Yen and Y-S. Lee. Under-sampling approaches for improving prediction of the minority class in an imbalanced dataset. In *Intelligent Control and Automation*, volume 344, pages 731–740, 2006.

[38] N.J. Yuan, Y. Zhong, F. Zhang, X. Xie, C-Y. Lin, and Y. Rui. Who will reply to/retweet this tweet?: The dynamics of intimacy from online social interactions. In *WSDM*, pages 3–12, 2016.

[39] C. Zhai and J. Lafferty. A study of smoothing methods for language models applied to information retrieval. *ACM Trans. Inf. Syst.*, 22(2):179–214, 2004.