



2SRM: Learning social signals for predicting relevant search results

Ismail Badache

► To cite this version:

Ismail Badache. 2SRM: Learning social signals for predicting relevant search results. Web Intelligence and Agent Systems, 2020, 18 (1), pp.15-33. 10.3233/WEB-200426 . hal-02927196

HAL Id: hal-02927196

<https://hal.science/hal-02927196>

Submitted on 2 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

2SRM: Learning Social Signals for Predicting Relevant Search Results

Ismail Badache *

Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France

Aix Marseille Univ, INSPE, SFEREP, Marseille, France

Abstract. Search systems based on both professional meta-data (e.g., title, description, etc.) and social signals (e.g., like, comment, rating, etc.) from social networks is the trending topic in information retrieval (IR) field. This paper presents 2SRM (Social Signals Relevance Model), an approach of IR which takes into account social signals (users' actions) as an additional information to enhance a search. We hypothesize that these signals can play a role to estimate a priori social importance (relevance) of the resource (document). In this paper, we first study the impact of each such signal on retrieval performance. Next, some social properties such as *popularity*, *reputation* and *freshness* are quantified using several signals. The 2SRM combines the social relevance, estimated from these social signals and properties, with the conventional textual relevance. Finally, we investigate the effect of the social signals on the retrieval effectiveness using state-of-the-art learning approaches. In order to identify the most effective signals, we adopt feature selection algorithms and the correlation between the signals. We evaluated the effectiveness of our approach on both IMDb (Internet Movie Database) and SBS (Social Book Search) datasets containing movies and books resources and their social characteristics collected from several social networks. Our experimental results are statistically significant, and reveal that incorporating social signals in retrieval model is a promising approach for improving the retrieval performance.

Keywords: Social Signals, Social Properties, Social Information Retrieval, Criteria Evaluation, Learning Approaches

1. Introduction

Before web 2.0, user interactions were limited to creation of links from one website to another one [50]. Nowadays, social web has completely changed the manner in which people communicate and share information on the web. It allows users to interact and produce large masses of social signals. On Facebook, the Like and Share Buttons are viewed across almost 10 million websites daily¹. On Twitter, second most popular social network after Facebook, thanks to its functionalities of *tweet* and *retweet* more than 150 million tweets were published just for the 2012 Olympics games². Other types of functions such as *endorsement*, *share*, *comment* and *rating* allow users to interact with web resources. Through these social actions, web re-

sources could become popular by accumulating the counts that people share such information, facilitate and help, users access novel information in convenient manner.

While we witness some recent moves from big players towards a more social information retrieval (such as Google and Bing expansion of results with those *liked* by the users' Facebook friends), the ways search engines and/or web 2.0 applications exploit social signals (if they ever do) are usually not disclosed. This paper describes an approach that exploits social networks or involve a collective intelligence process to help the user satisfy an information need. Particularly, we focus on exploiting social signals to estimate the social importance (relevance) of the resource to a given query. The main research questions addressed in this paper are the following:

* Corresponding author. E-mail: Ismail.Badache@lis-lab.fr

¹<https://zephoria.com/top-15-valuable-facebook-statistics/>

²<https://blog.twitter.com/2012/olympic-and-twitter-records>

1. Can these social signals help the search systems for guiding its users to reach a better quality or more relevant content?
2. How effective is each individual signals for ranking resources for a given query? What are the ranking correlations created by these social signals?
3. How to combine these social signals in form of social properties? What are the most useful of them to take into account in search model?

Note that we have already investigated the impact of social signals on search effectiveness using machine learning approaches in paper [11]. It only described some preliminary results and it did not deeply evaluate and analyze the results. This paper extends significantly our previous work in the following main additional aspects:

1. presenting a more complete state-of-the-art on social signals,
2. exploiting additional signals (e.g., tag, rating and its freshness, etc),
3. conducting new and extensive experiments on two standard datasets, namely INEX SBS (Social Book Search) and INEX IMDb (Internet Movie Database),

The remainder of this paper is organized as follows: Section 2 presents some related work and the background. Section 3 details our social IR approach. Then, Section 4, reports on the results of our experimental evaluation. Finally, Section 5 concludes this paper by announcing perspectives.

2. Background and Related Work

In this section, we report: (i) some background information about social signals and the existing search engines using these signals (e.g., *Google* and *Bing*); (ii) related work exploiting social data to estimate document relevance.

2.1. Social Signals

Social signals represent one of the most popular UGC (User Generated Content) on the web. Indeed, according to [4] web pages include buttons of different social networks where users can express whether they support, recommend or dislike content (text, image, video, etc). These buttons which describe actions

being of social activities (e.g., *like*, *share*, *+1*) are related to specific social networks (e.g., Facebook and Google+) with counters indicating the interaction rate with the web resource (see Figure 1).



Fig. 1. Example of a resource with social signals buttons

A social signal is a measure of the activity on social media. It is a social interaction of a real person with a resource on the web through the functionalities offered by social networks. As with backlinks³, social actions (e.g., *rating*, *like*, *like*, *+1*) can be interpreted as an approval for the resource, which can help to improve its ranking in search engines. As for political information, it is information related to government or political news, for example information about the election of regional heads. Social media enable users to get and share political information in social media and also giving opinions that can be done by pressing *like*, *dislike* buttons or *commenting* on information [43].

2.1.1. Types of Social Signals

In general, each social network uses its own social signals with different operating rules. Table 1 summarizes the most popular signals on social networks.

2.2. Social Signals and Search Engines

Despite the lack of clear consensus on the exact relationship between social signals and famous search engines (e.g., *Google* and *Bing*), there are many reasons why the signals can not be ignored. Rather than seeing

³backlinks or inbound link is a hyperlink to a site or web page.

Table 1
List of different social signals types

Type	Example	Social network
Vote	Like +1	Facebook, LinkedIn, Google+, StumbleUpon
Message	Tweet Post	Facebook, Google+, LinkedIn, Twitter
Share	Share Re-tweet	Google+, Twitter, Buffer, Facebook, LinkedIn
Tag	Bookmark Pin	Delicious, Diigo, Digg Pinterest
Comment	Comment Reply	Facebook, Google+, LinkedIn, Twitter
Emotion	Love, Haha, Wow Sad, Angry	Facebook
Event Reaction	Pride (Pride month) Thankful (Mother's day)	Facebook
Relation	Followers Friends	Facebook, Twitter

the signals and ranking of results by the search engines as two distinct components. It is useful to consider them as interconnected processes working towards the overall goal: increasing online visibility. Since inception of Facebook or other social networks, social signals also became an important information for SEO (Search Engine Optimization) [39]. They provide information about social interactivity, social behavior and social relations⁴. The correlation between social signal and the ranking position of a URL is extremely high. This is valid for all social networks covered by SISTRIX⁵ Toolbox. The works of Lewoniewski *et al.* [39] lead to the assumption that the results of social signals also correlate with the quality of Wikipedia articles. Therefore, social signals can be an indicator of the relevance of web resources.

Google is still mysterious about the way to exploit social signals to rank its search results, but some studies conducted annually since 2016, searchmetrics⁶ showed that it exists a high correlation between social signals and the rankings provided by search engines such as Google. However, the degree to which social signals play a role SEO is unclear. John Mueller (Webmaster Trends Analyst at Google) said : *"Do social media signals have an impact on organic rankings in Google? Not directly. No. So it's not that there's any kind of a ranking effect there. To a large part social networks also have a nofollow on the links that they provide when they post this content, so it's not the case*

that would give you any kind of a ranking boost there. What you do sometimes see however is that the social posts show up in the search results."⁷

Although Google does not have partnership with Facebook, it still has access to public data from Facebook and may use some of them to better understand the *popularity* of web pages. In 2015, Google signed an agreement with Twitter to index tweets in real time, allowing them to be searchable. Free access to Twitter database means that all the information on Twitter are available for Google automatically without the use of robots crawling. Google algorithms also focus on Twitter profiles who tweet and retweet content, but how do it remains a black box.

In addition, it is no secret that Google gives weight to its own social network, Google+. The content and interactions on Google+ are known to have a positive impact on the ranking of its results. The 2016 Searchmetrics study showed that +1 is strongly correlated with the ranking of Google search results, compared to other well-known criteria such as Facebook signal and keywords frequency.

Bing, the second biggest search engine after Google, is explicit in its use of signals such as *tweet* and Facebook *like* as well as other social signals as ranking factors [2,37,45]. Bing has partnered with Facebook in social search [51]. Bing algorithms focuses on the social media content, links, *popularity* from various social networks that are considered as important factors by Bing to define the ranking of results.

Bing is a notable example of exploiting images and information (posts, signals) from social media to provide the richest results for users [45]. The social media activity is also presented on Bing results pages much more visible than other search engines.

Social content such as *tweets*, *pins* (coming from Pinterest) and Facebook data, containing relevant keywords, are often integrated into the Bing search results. Thus, publishing visual contents (image, video, etc.) on the social networks is an excellent way to increase the visibility on Bing. Also, Bing is implementing a service called *Social Sidebar* that exploits Facebook to enhance a search [28]. It is a third column of the results page that allows users connected to *comment* and *like* the relevant results from Facebook without leaving the search page. This service is functional only in USA.

The popularity of UGC, particularly in the context of social media has given birth to many new problems

⁴https://en.ryte.com/wiki/Category:Search_Engine_Optimization

⁵<https://www.sistrix.fr/>

⁶<https://www.searchmetrics.com/knowledge-base/ranking-factors-infographic-2016/>

⁷<https://www.youtube.com/watch?v=WszvyRune14#t=1192>

in information retrieval. Specifically, how to exploit these social content in favor of IR is an open question, which gave birth to a new field in IR, Social Information Retrieval (SIR) [6,18].

2.3. Social Signals Approaches

Some approaches of social search rely on exploiting tags. Abel *et al.* [1] and Hotho *et al.* [31] proposed different algorithms based on folksonomies. By analogy with PageRank, FolkRank is based on relation between tag, resource and user, whereas with GFolkRank a resource group is identified by a unique tag in the context of the resource group. In the same context Bao *et al.* [17] proposed two algorithms, *SocialPageRank* and *SocialSimRank*, together with Yanbe *et al.* [50] suggest SBRank algorithm. These algorithms are motivated by the report that there is a strong interdependence between the popularity of users, tags and resources in a folksonomy. They focus on collective social search and doesn't respect different engagement types or different trust levels. *SocialSimRank* calculates similarity between two tags of folksonomy and declares that similar tags are usually assigned to similar resources. In addition, tags are sometimes more reliable than metadata provided by the content producer. However, single tag can hardly cover an entire topic and is more ambiguous for the user that a contextual sentence.

Furthermore, there are several recent works that focus on how to improve information retrieval (IR) effectiveness by exploiting the users' actions and their underlying social network. Chelaru *et al.* [22] studied the impact of signals (*like*, *dislike*, *comment*) 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 improve the ranking of search results for 48% of queries. They used feature selection algorithms and learning to rank algorithms. Karweg *et al.* [35] proposed an approach combining topical score and social score based on two factors: first, user engagement intensity quantifies the effort a user has made during an interaction with document, measured by the number of *clicks*, number of *votes*, number of *records* and *recommendation*, secondly, trust degree measured from social graph for each user according to his popularity, using PageRank algorithm. They have found that social results are available for most queries and usually lead to more satisfying results. Similarly, in [36,37] Khodaei *et al.* proposed a ranking approach based on several so-

cial factors including relationships between document owners and querying user, importance of each user and users' actions (playcount: number of times a user listens to a track on last.fm) performed on web documents. They have conducted an extensive experiments set on last.fm dataset. They showed a significant improvement for socio-textual ranking compared to the textual only and social only approaches. On Twitter, Hong *et al.* [30] used *retweets* as a measure of popularity and apply machine learning techniques to predict how often new messages will be retweeted. They exploited different features, the content of messages, temporal information, metadata of messages and users, and the user's social graph. However, banal tweets (e.g., rumors, without interest) can be very popular, such as those concerning celebrities, who generally have a large number of followers. Chan *et al.* [21] proposed a system called PostScholar, a service that augments the results returned by Google Scholar, a search engine for academic citations. PostScholar detects the Twitter activity related to an article and displays that information on the search results page returned by Google Scholar. An additional hyperlink appears in the results for each article that has Twitter activity associated with it (the number of tweets found for that article, the date of the most recent tweet). These tweets are sorted according to their sentiments scores. Albishre *et al.* [3] proposed an innovative mechanism to automatically select useful feedback documents using a topic modeling technique to improve the effectiveness of pseudo-relevance feedback models. The main idea behind their proposed model is to discover the latent topics in the top-ranked documents that allow for the exploitation of the correlation between terms in relevant topics. To capture discriminating terms for query expansion, they incorporated topical features into a relevance model that focuses on the temporal information in the selected set of documents. Experimental results on TREC 2011-2013 microblog datasets illustrated that the proposed model significantly outperforms all state-of-the-art baseline models.

Finally, there are other studies initiated by Microsoft Bing researchers [41,45] that show the usefulness of different social contents generated by the network of user's friends on Facebook. Pantel *et al.* [42] studied the leverage of social annotation on the quality of search results. They observed that social annotations can benefit web search in two aspects: 1) the annotations are usually good summaries of corresponding web pages, 2) the number of annotations indicates the popularity of web pages. Hecht *et al.* [28] presented

a system called SearchBuddies based on any social information around the user and especially what his friends *liked* and *shared* as web page, Facebook pages. Gou *et al.* [24] proposed a ranking approach taking into account document content and similarity between user and documents user owner in social network. They used a multi-level algorithm to measure the similarity between actors. Experimental results based on YouTube data show that compared with tf-idf algorithm, SNDocRank method returns more relevant documents. According to these results, a user can enhance search by joining a larger social networks, having more friends, and connecting larger communities.

In this paper, our goal aims to exploit social signals for improving accuracy and relevance of conventional textual web search. We exploit various signals extracted from different social networks. In addition, instead of considering social features separately as done in the previous works, we propose to combine them to measure specific social properties, namely the popularity and the reputation of a resource. We also attempt to measure the impact of the freshness of the signal on the performance of IR system. Unlike our previous work on social signals [5,7,8,10,12,13,14,15,16], in this paper, the proposed approach is completely supervised by exploiting social signals collected from different social networks as criteria of relevance. It is evaluated on different types of standard test data (INEX Social Book Search and INEX Internet Movie Database). In the following, we present some previous work in connection with this work but are different in regard to the solution proposed in this paper.

In [5] we presented our first basic preliminary study in the form of a poster which proposes a problematic on the users' information needs coming from the social web, this problem is related to what is called social search or social information retrieval. In [10,12] we have used some features such as comments and likes for ranking web resources, and addressing the impact of number of shares and likes in search relevance. The use of temporal features is inspired by the use of features of this class (e.g., Age of the resource and the signal) in specific domains, such as IMDb movies [13,14]; In these papers, we are particularly interested in: first, showing the impact of signals diversity associated to a resource on information retrieval performance; second, studying the influence of their social networks origin on their quality. We have proposed to model these social features as prior that we integrate into an unsupervised language model. In [15,16] our objective is to study the impact of the new emotional

social signals, called Facebook reactions (love, haha, angry, wow, sad) in the retrieval. These reactions allow users to express more nuanced emotions compared to classic signals (e.g. like, share). First, we have analyzed these reactions and showed how users use these signals to interact with posts. Second, we have evaluated the impact of each such reaction in the retrieval, by comparing them to both the textual model without social features and the first classical signal (like-based model). Similarly to the modeling signals in [13,14], these social features are modeled as document prior and are integrated into a language model. We have conducted a series of experiments on IMDb dataset. In [8,9] our goal is to show how these social traces can play a vital role in improving Arabic Facebook search. Firstly, we have identified polarities (positive or negative) carried by the textual signals (e.g. comments) and non-textual ones (e.g. the reactions love and sad) for a given Facebook post. Therefore, the polarity of each comment expressed on a given Facebook post, is estimated based on a neural sentiment model in Arabic language. We note that sentiment analysis of social content is a complex task [20]. Secondly, we have grouped signals according to their complementarity using features selection algorithms. Thirdly, we have applied learning to rank (LTR) algorithms to re-rank Facebook search results based on the selected groups of signals. Finally, experiments are carried out on 13,500 Facebook posts, collected from 45 topics in Arabic language. Experiments results reveal that Random Forests combined with ReliefFAttributeEval (RLF) was the most effective LTR approach for this task. In [7] we have conducted an exploratory study in the impact of users' traces on Arabic and English Facebook search. In general, during all these years of research, our findings reveal that incorporating social features is a promising approach for improving the IR ranking performance in Arabic and English languages.

3. 2SRM: Social Signals Relevance Model

Our IR approach, named 2SRM, consists of exploiting social signals to define social properties to take into account in retrieval model. We associate to each web resource a social relevance estimated based on these social features (signals and properties). The social relevance score is then combined with a classical topical relevance score (see Figure 2).

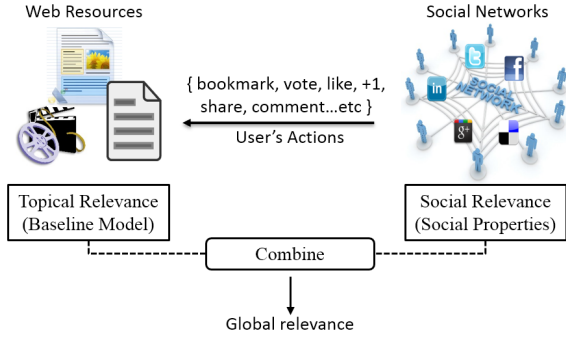


Fig. 2. A modular approach for Social IR

3.1. Notation

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

Resources. We consider a collection $R = \{r_1, r_2, \dots, r_n\}$ of n resources. Each resource r can be a web page, video or other type of web resources. A resource r can be represented both as a set of textual keywords $r_w = \{r_{w_1}, r_{w_2}, \dots, r_{w_n}\}$ and as a set of social actions $A = \{a_1, a_2, \dots, a_m\}$ performed on this resource.

Actions. There is a set $A = \{a_1, a_2, \dots, a_m\}$ of m actions that users can perform on the resources. These actions represent the relation between users $U = \{u_1, u_2, \dots, u_l\}$ and resources R . For instance, on Facebook, users can perform the following actions on resources: *publish*, *like*, *share* or *comment*.

Time. It represents the history of each action, let $T_{a_i} = \{t_{1,a_i}, t_{2,a_i}, \dots, t_{k,a_i}\}$ a set of k moments (date) at which each action a_i was produced. A moment t represents the datetime for each action a of the same type.

Social networks. There is a set $G = \{g_1, g_2, \dots, g_z\}$ of z social networks. Each specific social network contains specific social actions a performed on resources.

3.2. Formalization of Our Social Search Model

By analyzing various types of social actions (signals) through many social networks, we define three social properties that are detailed below:

Popularity P. Is a social phenomenon which indicates which is the most known among the public. Thanks to the influence of peers, target resources progress quickly in terms of its pervasive in the society. Therefore, the web resource popularity can be estimated according to the rate of sharing this resource between the users through social actions. We assume

that the popularity comes from users' activities on social networks, i.e. A resource is said popular if it was *shared* and *commented* by several users in several social networks, to the point where it becomes very known to general public.

Reputation R. The resource popularity does not reflect its good or bad reputation. Resource reputation is an opinion on this resource, we believe that the estimation of this property can be calculated based on social activities that have positive meaning such as Facebook *like* or *marking* resource as favorites on Delicious. Indeed, resource reputation depends on the degree of users' appreciation on social networks.

In a summary, we assume that some social actions are more suitable to evaluate popularity of a resource and others are more related to reputation. Therefore, we associate to each of these properties a score calculated by a simple counting (normalized using min-max) of the number of associated actions. The general formula is the following:

$$f_x(r, G) = \sum_{i=1, a_i^x \in A}^m \text{Count}(a_i^x, r, G) \quad (1)$$

Formula 1 is normalized as follows:

$$f_x(r, G)_{\text{Norm}} = \frac{f_x(r, G) - \text{Min}(f_x(r, G))}{\text{Max}(f_x(r, G)) - \text{Min}(f_x(r, G))} \quad (2)$$

Where:

- $\text{Count}(a_i^x, r, G)$ represents number of occurrence of specific action a_i^x performed on a resource in relation to a specific social network. $x = \{P, R\}_{\text{social}}$ (with P : Popularity, R : Reputation).
- $f_x(r, G)$ arithmetic function that represents the linear combination of m social actions that quantify each x social properties (rate of interaction through the social signals).
- $\text{Min}(f_x(r, G)), \text{Max}(f_x(r, G))$ represent the minimum and maximum value for f_x . $f_x(r, G)_{\text{Norm}}$ represents the f_x min-max normalization.

In addition to a simple counting of social actions, we propose to consider the time associated with the signal. We assume that the resource associated with fresh (recent) signals should be promoted.

Freshness F. Is an important relevance factor, exploited by several search engines. The information freshness is often measured in relation to its publica-

tion date, but we cannot say that information is necessarily obsolete because it was published two years ago. Taking an example of a resource published in September 2001, carrying an information about the attack on "World Trade Center", in 2019, the same resource was discussed on social networks through different social signals. We assume that a resource is fresh if recent social signals were associated with it. For that purpose, we define freshness as follow: "a date of each social action (e.g., date of comment, date of share) performed on a resource on social networks can be exploited to measure the recency of these social actions, hence the freshness of information". Its formula is given as follows:

$$f_F(r, G) = \frac{1}{\frac{1}{m} \sum_{i=1}^m \left(\frac{1}{k} \sum_{j=1}^k Time(t_{j,a_i}, r, G) \right)} \quad (3)$$

Where:

- $Time(t_{j,a_i}, r, G)$ calculating the time duration (recency) between current time and action time t_{j,a_i} of the same type for a resource r . We notice that for each action, its time is initialized to: 01-01-1970 00:00:00.
- $f_F(r, G)$ freshness function that represents the inverse average of $Time(t_{j,a_i}, r, G)$ values for a resource r .
- $\frac{1}{k} \sum_{j=1}^k Time(t_{j,a_i}, r, G)$ represents the time duration average related to all k action time t_{j,a_i} of the same type.
- $\frac{1}{m} \sum_{i=1}^m \left(\frac{1}{k} \sum_{j=1}^k Time(t_{j,a_i}, r, G) \right)$ represents the average value of time duration averages related to all m types of action on the resource.

Social score. Regarding the social score $Rel_S(q, r, G)$, we specify that this score takes into account these social properties, which are in the form of three normalized factors that are combined linearly by the following formula:

$$Rel_S(q, r, G) = \delta \cdot f_R(r, G) + \lambda \cdot f_P(r, G) + \beta \cdot f_F(r, G) \quad (4)$$

Where β, λ, δ are parameters assigning relative weights to each social properties, $\forall(\beta, \lambda, \delta) \in [0, 1]^3, \beta + \lambda + \delta = 1$.

Example : How to calculate Rel_S ?

Let $R = \{r_1, r_2, r_3\}$ a collection of 3 resources.

Let $A = \{a_1, a_2, a_3, a_4\}$ of 4 actions (signals) that users can perform on the resources, where : $a_1 = like$, $a_2 = bookmark$, $a_3 = tweet$ and $a_4 = share$. We hypothesized that reputation R is quantified using a_1 and a_2 , and popularity P is quantified using a_3 and a_4 .

Table 2 illustrates, through an example of data, the different steps to calculate the final social score.

Table 2
Calculation example of Rel_S

Functions / Resources	r_1	r_2	r_3
Calculation of Reputation R			
$Count(a_1^R, r, G)$	100	11	22
$Count(a_2^R, r, G)$	77	200	100
eq. 1 : $f_R(r, G)$	177	211	122
eq. 2 : $f_R(r, G)_{Norm}$	0.28	1	0
Calculation of Popularity P			
$Count(a_3^P, r, G)$	111	17	35
$Count(a_4^P, r, G)$	18	250	112
eq. 1 : $f_P(r, G)$	129	267	147
eq. 2 : $f_P(r, G)_{Norm}$	0	1	0.13
Calculation of Freshness F			
$\left(\frac{1}{k} \sum_{j=1}^k Time(t_{j,a_1}, r, G) \right)$	883	39	189
$\left(\frac{1}{k} \sum_{j=1}^k Time(t_{j,a_2}, r, G) \right)$	128	50	111
$\left(\frac{1}{k} \sum_{j=1}^k Time(t_{j,a_3}, r, G) \right)$	1519	18	372
$\left(\frac{1}{k} \sum_{j=1}^k Time(t_{j,a_4}, r, G) \right)$	350	31	570
$\frac{1}{4} \sum_{i=1}^4 \left(\frac{1}{k} \sum_{j=1}^k Time(t_{j,a_i}, r, G) \right)$	720	34.5	310.5
eq. 3 : $f_F(r, G)$	0.001	0.03	0.003
Social score Rel_S : e.g. $\delta = 0.6, \lambda = 0.2, \beta = 0.2$			
eq. 4 : $Rel_S(q, r, G)$	0.224	1	0.026
Ranking of resources	2	1	3

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 first evaluated the impact of social signals, taken separately and when they are combined as properties (popularity, reputation and freshness). Secondly, we study the effectiveness of each social signal using machine learning with selection attributes algorithms. We compared our approach which takes into account social signals, with the baseline formed by only a textual model. Our main goals in these experiments are:

1. to evaluate the impact of signals taken separately and grouped (properties),
2. to evaluate the effectiveness of signals using machine learning techniques.

4.1. Description of INEX Datasets (IMDb and SBS)

We used the collections SBS⁸ and IMDb⁹ 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 (see table 3 and 5). For each document, we collected specific social signals via their corresponding API of 6 social networks listed in Table 7. 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.

Table 4 presents statistics on the number of signals in the 1000 documents returned by each topic (30 IMDb topics). According to the averages of signal numbers in the documents, we note that the density of Facebook signals (in average: 85.8 like, 94.1 share and 98.4 comment) is very high compared to other signals (in average: 2.5 +1, 0.9 bookmark, 17.2 tweet and 1.4 share (LinkedIn)).

Table 6 presents statistics on the number of signals in the 1000 documents returned by each topic (208 SBS topics). We note that the density of Facebook signals is very high compared to Amazon/LibraryThing signals but the total number of *rating* and *review* is much higher compared to other signals.

4.2. Quantifying Social Properties

Table 7 presents the properties that we want to take into account in our retrieval model. In order to quantify these social properties, we associate them with the corresponding social signals. Specific social signals (actions) have been associated with each property depending on their nature and meaning. In Table 7, we note that the social signals that quantify reputation carry positive opinions, for example, *bookmark* a resource link by a user on Delicious means that this resource has been added to his favourites list. Concerning *like* and *rating*, user clicks on these buttons to indicate that he has enjoyed the resource content. So the presence of these social signals in resource increases the degree

of resource reputation. The same applies for popularity, the exploited social signals to estimate it, let us know the position of this resource on the web in terms of trend and propagation. Finally, to quantify freshness the dates of the different actions are not available except the dates of each rating from SBS dataset.

4.3. Result of Linear Combination Study

We conducted experiments with models based only on the contents of documents, as well as approaches combining content and social data. Normalized formula (5) is the weighted sum of social relevance and topical relevance:

$$Rel(q, r, G) = \alpha \cdot Rel_T(q, r) + (1 - \alpha) \cdot Rel_S(q, r, G) \quad (5)$$

Where: $\alpha \in [0, 1]$ as a weighting parameter and $Rel_T(q, r)$ is the normalized score of topical relevance. We used BM25 [46] and Lucene Solr scoring¹⁰ as baselines models for our study. Lucene Solr scoring uses a combination of the Vector Space Model (VSM) and the Boolean model.

In this paper, we evaluate the contribution of each social signal/social property and the effect of their combination on relevance. We first select the best parameters α (see formula 5) and β, λ, δ (see formula 4) by applying J48 learning algorithm, then we compare our approach with baselines. We note that if $\alpha = 0$ only the social relevance is taken into account. Moreover, $\alpha = 1$ corresponds to the baselines textual models. The best values of parameters are the following: $\alpha \in [0.5, 0.6]$ with $\beta \in [0.1, 0.2]$, $\lambda \in [0.3, 0.5]$, $\delta \in [0.4, 0.6]$ for P@10 and P@20.

Tables 8 and 9 summarize the results of precision (P@10 and P@20), nDCG and MAP for IMDb and SBS datasets, respectively. We evaluated different configurations, by taking into account social signals individually and their combination in the form of social properties. the "Freshness" configuration is only estimated for the SBS collection where we have the appearance time of each rating on a document. In order to check the significance of the results, we performed the Student test [23] and attached * (strong significance against BM25) and ** (very strong significance against BM25) to the performance number of each row in the tables 8 and 9 when the p-value < 0.05 and p-value < 0.01 confidence level, respectively.

⁸<http://social-book-search.humanities.uva.nl/#/data>

⁹<https://inex.mmci.uni-saarland.de/tracks/dc/2011/>

¹⁰<http://lucene.apache.org/solr/>

Table 3
List of the different IMDb document metadata fields

Field	Description	Status	Field	Description	Status
<i>ID</i>	Identifying the film (document)	-	<i>Director</i>	Director of the film project	Indexed
<i>Title</i>	Film's title	Indexed	<i>Writer</i>	Writers and writers of the film	Indexed
<i>Year</i>	Year of the film release	Indexed	<i>Actors</i>	Main actors of the film	Indexed
<i>Rated</i>	Film classification by content type	-	<i>Plot</i>	Text summary of the film	Indexed
<i>Released</i>	Date of making the film	Indexed	<i>Poster</i>	URL of the link poster	-
<i>Runtime</i>	Length of the film	Indexed	<i>url</i>	URL of the web source document	-
<i>Genre</i>	Film genre (Action, Drama, etc.)	Indexed	<i>UGC</i>	Social signals recovered	-

Table 4
Statistics on the number of signals in the IMDb documents returned by 30 topics

Network	Facebook			Google+	Delicious	Twitter	LinkedIn
Signal	Like	Share	Comment	+1	Bookmark	Tweet	Share
Minimum	0	0	0	0	0	0	0
Maximum	76842	43918	62281	1475	986	12223	299880
Total	2478498	2718918	2845169	73392	26143	499232	42787
Average	85.8027	94.1258	98.4964	2.5407	0.9050	17.2830	1.4812

Table 5
List of the different SBS document fields

Field	Status	Field	Status	Field	Status	Field	Status
book	Index	similarproducts	-	title	Index	imagecategory	-
dimensions	-	tags	Index	edition	Index	name	Index
reviews	Index	isbn	Index	dewey	-	role	Index
editorialreviews	Index	ean	Index	creator	Index	blurber	-
images	-	binding	Index	review	Index	dedication	Index
creators	Index	label	Index	rating	-	epigraph	Index
blurbers	-	listprice	-	authorid	-	firstwordsitem	Index
dedications	Index	manufacturer	Index	totalvotes	-	lastwordsitem	-
epigraphs	Index	numberofpages	-	helpfulvotes	-	quotation	Index
firstwords	Index	publisher	Index	date	Index	seriesitem	-
lastwords	Index	height	-	summary	Index	award	-
quotations	Index	width	-	editorialreview	Index	browseNode	-
series	Index	length	-	content	Index	character	-
awards	-	weight	-	source	Index	place	Index

Table 6
Statistics on the number of signals in the SBS documents returned by 208 topics

Network	Facebook			Amazon		
Signal	Like	Share	Comment	Review	Rating	Tag
Minimum	0	0	0	0	0	0
Maximum	7213	5892	5975	3378	3378	277
Total	4760698	5222491	5465011	23856118	23856118	1811968
Average	41.3082	45.3152	47.4195	206.9981	206.9981	15.7223

Table 7

Exploited social signals in the quantification

Property	c_i	Social Signal	Source	Dataset
Popularity	c_1	Number of <i>Comment</i>	Facebook	SBS, IMDb
	c_2	Number of <i>Tweet</i>	Twitter	IMDb
	c_3	Number of <i>share(LIn)</i>	LinkedIn	IMDb
	c_4	Number of <i>share</i>	Facebook	SBS, IMDb
	c_5	Number of <i>Review</i>	Amazon/LibraryThing	SBS
	c_6	Number of <i>Tags</i>	Amazon/LibraryThing	SBS
Reputation	c_7	Number of <i>Like</i>	Facebook	SBS, IMDb
	c_8	Number of <i>Mention +1</i>	Google+	IMDb
	c_9	Number of <i>Bookmark</i>	Delicious	IMDb
	c_{10}	Average of <i>Rating</i>	Amazon/LibraryThing	SBS
Freshness	c_{11}	<i>Dates of Ratings</i>	Amazon/LibraryThing	SBS

Table 8

Comparing social search effectiveness to Solr and BM25 (On IMDb)

	P@10	P@20	nDCG	MAP
BM25	0.2912	0.2276	0.3158	0.1733
Lucene Solr	0.2617	0.1951	0.2744	0.1590
Like	0.3963**	0.3142**	0.4873**	0.2339**
Share	0.4037**	0.3119**	0.4911**	0.2428**
Comment	0.3855**	0.3004**	0.4103**	0.2141**
Tweet	0.3488**	0.2613**	0.4096**	0.1882**
Mention+1	0.3206**	0.2467**	0.3627**	0.1809**
Share(LIn)	0.3177*	0.2411*	0.3509*	0.1739*
Bookmark	0.3358*	0.2395*	0.3517*	0.1730*
All Criteria	0.4225**	0.3702**	0.4974**	0.2701**
Reputation	0.4513**	0.3961**	0.5478**	0.2671**
Popularity	0.4167**	0.3622**	0.5510**	0.2612**
All Properties	0.4984**	0.4372**	0.5971**	0.2916**

Table 9

Comparing social search effectiveness to Solr and BM25 (On SBS)

	P@10	P@20	nDCG	MAP
BM25	0.0601	0.0517	0.1581	0.0540
Lucene Solr	0.0528	0.0487	0.1300	0.0463
Like	0.0768**	0.0600**	0.1776**	0.0652**
Share	0.0813**	0.0622**	0.1812**	0.0784**
Comment	0.0701**	0.0588**	0.1719**	0.0613**
Tag	0.0642*	0.0471*	0.1660*	0.0532*
Rating	0.0672**	0.0442**	0.1654**	0.0622**
All Criteria	0.0883**	0.0699**	0.1893**	0.0820**
Freshness	0.0853**	0.0644**	0.1816**	0.0795**
Reputation	0.0884**	0.0713**	0.1886**	0.0820**
Popularity	0.0876**	0.0692**	0.1865**	0.0801**
All Properties	0.0932**	0.0800**	0.1912**	0.0843**

We observe in all cases, with taking into account social features, the results are significantly better compared to textual models. Also, it is clear that combining social signals as properties (popularity and reputation) provides better results than when they are taken indi-

vidually. The results show that *reputation* configuration provides better results than *popularity*. The freshness in our study is related to the recency of actions associated with a resource. The resources that possess fresh signals are promoted in search results list. The overall combination of social properties provides the best results. According to Student test, majority of the results show a strong statistically significant improvement.

In general, experimental results reflect the effectiveness of social signals on search task. More specifically, the results show that the way we have combined social signals to quantify different properties is effective to improve precision and nDCG. Therefore, combination of *freshness* with *popularity* and *reputation* (named "All Properties" in tables 8 and 9) provides the best improvement compared to a random combination of "All Criteria".

We can explain these results by the positive sense of *reputation* property quantified through the counting of signals such as *like*, *mention+1*, *bookmark* and *rating*, which means favourable and positive opinion for the resource judgment. Social networks urge users to *share*, *comment*, *evaluate* and *disseminate* the information on a large scale. These interactions allow us to draw conclusions about the social position and quality of these resources in social networks across their *popularity*, *reputation* and *freshness*. Therefore, we can also explain our results by the high rate of user's engagements on various social networks, which brings together more than a billion users, producing users' massive interactions "wisdom of crowds" with web resources through these social actions of different natures, often positives.

4.4. Identify the Most Effective Signals

In this section, we conducted a series of experiments in a supervised environment, using machine learning algorithms with the set of effective social signals identified in table 7. The aim is twofold: on the one hand we wondered whether the attribute selection really improves the results of a search. On the other hand, we intended to measure the performance of some learning algorithms in this type of classification.

4.4.1. Methodology

Removing the irrelevant and redundant features from the data helps to improve the performance of learning models. Following are the most well-known feature selection algorithms:

- **CfsSubsetEval:** CFS, short for "Correlation-based Feature Subset Selection", evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. Subsets of features that are highly correlated with the class while having low intercorrelation are preferred [26].
- **WrapperSubsetEval:** Evaluates attribute sets by using a learning scheme. Cross validation is used to estimate the accuracy of the learning scheme for a set of attributes [38].
- **ConsistencySubsetEval:** Evaluates the worth of a subset of attributes by the level of consistency in the class values when the training instances are projected onto the subset of attributes. Consistency of any subset can never be lower than that of the full set of attributes, hence the usual practice is to use this subset evaluator in conjunction with a Random or Exhaustive search which looks for the smallest subset with consistency equal to that of the full set of attributes [33].
- **FilteredSubsetEval:** Class for running an arbitrary subset evaluator on data that has been passed through an arbitrary filter (note: filters that alter the order or number of attributes are not allowed). Like the evaluator, the structure of the filter is based exclusively on the training data [27].
- **ChiSquaredAttributeEval:** Evaluates the worth of an attribute by computing the value of the chi-squared statistic with respect to the class [32].
- **FilteredAttributeEval:** Class for running an arbitrary attribute evaluator on data that has been passed through an arbitrary filter (note: filters that alter the order or number of attributes are not al-

lowed). Like the evaluator, the structure of the filter is based exclusively on the training data [27].

- **GainRatioAttributeEval:** Evaluates the worth of an attribute by measuring the gain ratio with respect to the class [27].

$$\text{GainR}(\text{Class}, \text{Attribute}) = \frac{(H(\text{Class}) - H(\text{Class}|\text{Attribute}))}{H(\text{Attribute})} \quad (6)$$

While: H specifies the entropy. Entropy is a measure of the uncertainty associated with a random variable.

- **InfoGainAttributeEval:** Evaluates the worth of an attribute by measuring the information gain with respect to the class [27].

$$\text{GainR}(\text{Class}, \text{Attribute}) = (H(\text{Class}) - H(\text{Class}|\text{Attribute})) \quad (7)$$

- **OneRAttributeEval:** OneR, short for "One Rule", is a simple classifier that generates one-level decision tree. OneR evaluates the worth of an attribute by using the OneR classifier. OneR classifier used crossvalidation to estimate the accuracy of the learning scheme for a set of attributes [29]. It combines the C4.5 decision tree and Gaussian distribution.
- **ReliefFAttributeEval:** Evaluates the worth of an attribute by repeatedly sampling an instance and considering the value of the given attribute for the nearest instance of the same and different class. Can operate on both discrete and continuous class data [40].
- **SVMAttributeEval:** Evaluates the worth of an attribute by using an SVM classifier. Attributes are ranked by the square of the weight assigned by the SVM. Attribute selection for multiclass problems is handled by ranking attributes for each class separately using a one-vs-all method and then "dealing" from the top of each pile to give a final ranking [25].
- **SymmetricalUncertAttributeEval:** Evaluates the worth of an attribute by measuring the symmetrical uncertainty with respect to the class [26].

$$\text{SymmU}(\text{Class}, \text{Attribute}) = \frac{2 * (H(\text{Class}) - H(\text{Class}|\text{Attribute}))}{H(\text{Class}) + H(\text{Attribute})} \quad (8)$$

In this study, we relied on algorithms for selecting attributes to determine the best social signals to exploit in the learning model. Feature selection Algorithms [27] aim to identify and eliminate as many irrelevant and redundant information as possible. We used Weka¹¹ for this experiment. It is a powerful open-source Java-based learning tool that brings together a large number of learning machines and algorithms for selecting attributes.

We proceeded as follows: the top 1000 documents for each topic from the two collections (30 IMDb topics and 208 SBS topics) were extracted using the default Lucene Solr model. Then, the scores of all criteria (social signals) are calculated for each resource. We identified relevant documents and irrelevant documents according to the Qrels. The resulting set contains 30000 documents for IMDb and 115248 documents for SBS, including:

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

We observed that this collection has an unbalanced relevance classes distribution. This occurs when there are many more elements in one class than in the other class of a training collection. In this case, a classifier usually tends to predict samples from the majority class and completely ignore the minority class [48]. For this reason, we applied an approach to sub-sampling (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.

We note that these algorithms operate differently, some return an importance ranking of attributes (e.g., *FilteredAttributeEval*), while others return the number of times that a given attribute has been selected by an algorithm in a cross-validation (e.g., *FilteredSubsetEval*). In the following, we present the results obtained by each selection algorithm (with and without considering the time) applied on INEX SBS and IMDb datasets. We note that we have used for each algorithm the default setting provided by Weka.

Our goal of this study is to determine the most important signals for IR task and verify if the results obtained previously (Table 8) are consistent. In our case, the selection algorithms are to give a score to each signal based on its significance towards relevance class of the document (relevant or irrelevant). For IMDb (see Table 10), we evaluated 7 criteria where $n = 7$, and for SBS (see Table 11) we evaluated 6 criteria. We applied a cross-validation for 5 iterations (5 cross-validation folds). Table 10 shows the social signals selected through attributes selection algorithms. We use two types of these algorithms: a) those using ranking methods to order the selected criteria (metric in the tables is [Rank]); and b) those using search methods that indicate how many times the criterion has been selected during the cross-validation task (metric in tables is [Folds Number]). A social signal strongly preferred by the selection algorithm is a well-ranked signal ($rank = 1$) and strongly selected ($folds\ number = 5$).

According to Table 10, we remark that the Facebook signals c_1 : *Comment*, c_4 : *Share* and c_7 : *Like* are the most selected and highly ranked compared to other signals. Signals c_2 : *Tweet* and c_8 : *+1* are also favored by the attributes selection algorithms, except the algorithm *WrapperSubsetEval* that selected them only one time among 5 iterations, and the algorithms *ReliefFAttributeEval*, *SVMAttributeEval* that ranked c_2 : *Tweet* in 7 and 6, respectively. Finally, the most weakest and most disadvantaged signals are c_3 : *Share(LIn)* and c_9 : *Bookmark*. These results are correlated with the obtained results in Table 8.

Similarly, on SBS dataset, table 11 shows that Facebook signals c_1 : *Comment*, c_4 : *Share* and c_7 : *Like* are always highly ranked and often validated during the 5 iterations of cross-validation. In addition, the signal c_{11} : rating biased by its freshness is strongly favored (average ranks = 2 and average selection = 5). This last result proves the interest of the freshness that we propose, i.e. resources with fresh signals should be boosted in the top list of search results. The signals c_6 : *Tag* and c_{10} : average of rating (without consider-

¹¹<http://www.cs.waikato.ac.nz/ml>

Table 10
Selected signals with attribute selection algorithms (IMDb dataset)

Algorithm	Metric	c_1^{++}	c_2^+	c_3	c_4^{++}	c_7^{++}	c_8^+	c_9
CfsSubsetEval	[folds Number]	5	5	-	5	5	2	-
WrapperSubsetEval	[folds Number]	1	1	1	4	5	1	3
ConsistencySubsetEval	[folds Number]	5	5	5	5	5	5	4
FilteredSubsetEval	[folds Number]	5	5	-	5	5	2	-
ChiSquaredAttributeEval	[Rank]	3	4	6	1	2	5	7
FilteredAttributeEval	[Rank]	3	4	6	1	2	5	7
GainRatioAttributeEval	[Rank]	1	4	7	2	3	5	6
InfoGainAttributeEval	[Rank]	3	4	6	1	2	5	7
OneRAttributeEval	[Rank]	2	4	6	3	1	5	7
ReliefFAttributeEval	[Rank]	3	7	5	1	2	4	6
SVMAttributeEval	[Rank]	5	6	2	1	4	3	7
SymmetricalUncertEval	[Rank]	1	4	6	2	3	5	7
Count		12	12	10	12	12	12	10

Table 11
Selected signals with attribute selection algorithms (SBS dataset)

Algorithms	Metric	c_1^{++}	c_4^{++}	c_6^+	c_7^{++}	c_{10}^+	c_{11}^{++}
CfsSubsetEval	[folds number]	5	5	-	5	-	5
WrapperSubsetEval	[folds number]	1	5	2	4	4	4
ConsistencySubsetEval	[folds number]	4	5	3	5	4	5
FilteredSubsetEval	[folds number]	4	5	-	5	-	5
ChiSquaredAttributeEval	[rank]	4	1	6	2	5	3
FilteredAttributeEval	[rank]	4	1	6	2	5	3
GainRatioAttributeEval	[rank]	3	1	6	4	5	2
InfoGainAttributeEval	[rank]	4	2	6	3	5	1
OneRAttributeEval	[rank]	4	2	6	3	5	1
ReliefFAttributeEval	[rank]	4	1	6	3	5	2
SVMAttributeEval	[rank]	4	1	6	3	5	2
SymmetricalUncertEval	[rank]	4	2	6	3	5	1
Count		12	12	10	12	10	12

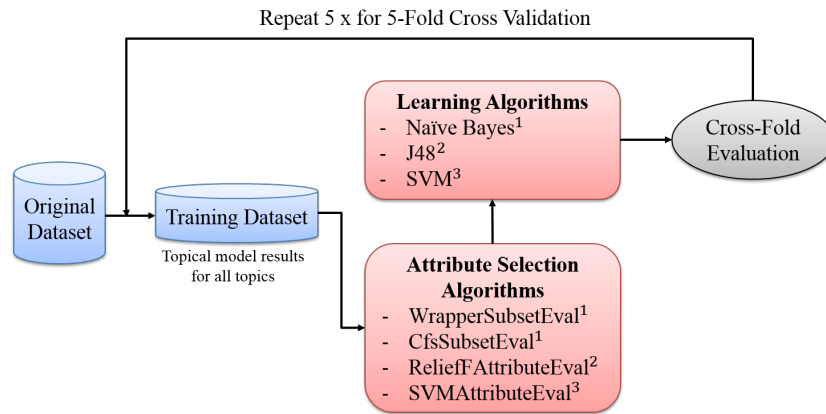


Fig. 3. Machine Learning Process

ing its creation dates) are weakly preferred (favored) by the selection algorithms, with average ranks of 6, 5, respectively, and the average selection of 1.25, 2, respectively. By comparing these results with the results listed in Table 9, we noticed that the same signals highlighted by the linear combination method are highlighted by the learning techniques.

4.5. Learning Signals for Predicting Relevance

We also conducted a series of experiments exploiting these signals in supervised approaches based on learning models. We used the results returned by Lucene Solr using all the queries from the two INEX collections (IMDb and SBS), each separately, as training sets. We then used three learning algorithms, this choice being explained by the fact that they often showed their effectiveness in IR by exploiting criteria of relevance: SVM [34], J48 (C4.5 implementation) [44] and Naive Bayes [49]. The input of each algorithm is a vector of social signals, either all the signals or just the signals selected by a precise selection algorithm. Each signal is represented by its quantity in the documents. Learning algorithms predict the relevance class for each document (relevant or irrelevant). We applied a cross validation for 5 iterations (5 cross-validation folds). Figure 3 illustrates the learning process we used for evaluating social signals.

We recall that the phase of attribute selection algorithms has highlighted two sets of signals:

1. in the case of the CfsSubsetEval and FilteredSubsetEval algorithms, the selected signals are c_1 : comment, c_4 : share, c_7 : like and c_{11} : rating (with considering its time) for SBS and c_1 : comment, c_2 : tweet, c_4 : share, c_7 : like, c_8 : +1 and c_{10} : rating (without considering its time) for IMDb.
2. in the case of the other selection algorithms, all the social signals studied on the two collections are selected: c_1 , c_4 , c_6 , c_7 , c_{10} and c_{11} for the SBS collection and c_1 , c_2 , c_3 , c_4 , c_7 , c_8 and c_9 for the IMDb collection.

The question at this stage relates to the specification of the input signal vector for the learning algorithms, either we take all the signals, or we keep only those selected by the attribute selection algorithms. In this case, with which learning algorithms these will be combined.

In order to take into account the signals chosen by the selection algorithms in learning models, we relied on the work of Hall and Holmes [27].

Hall and Holmes [27] have studied the effectiveness of some attribute selection techniques by confronting them with learning techniques. Since the performance of the factors differs from one learning technique to another, they have identified the best attribute selection techniques to find the best performing factors according to the learning techniques to be used.

Based on their study, we used the same pairs of learning techniques and attribute selection techniques:

- signals selected by WrapperSubsetEval (WRP) are learned by Naive Bayes.
- signals selected by CfsSubsetEval (CFS) are learned by Naive Bayes.
- signals selected by ReliefFAttributeEval (RLF) are learned by J48.
- signals selected by SVMAttributeEval (SVM) are learned by SVM.

Table 12
Machine learning results (P@20) on IMDb

Classifiers	Attribute selection criteria	All criteria
NaiveBayes	0.4927 (CFS)	0.4802
	0.4802 (WRP)	
SVM	0.4874 (SVM)	0.4874
J48	0.5562 (RLF)	0.5562

Table 13
Machine learning results (P@20) on SBS

Classifiers	Attribute selection criteria	All criteria
NaiveBayes	0.1223 (CFS)	0.1100
	0.1100 (WRP)	
SVM	0.1114 (SVM)	0.1114
J48	0.1301 (RLF)	0.1301

Tables 12 and 13 present the results of the three learning algorithms of signals that emerged from the study using attribute selection techniques. We find that only the CFS algorithm confirms the hypothesis put forward by Hall and Holmes. Indeed, it is the only one for which the results obtained with the selection of attributes, are 0.4927 (IMDb) and 0.1223 (SBS), exceed the use of all the signals, 0.4802 (IMDb) and 0.1100 (SBS). We have shown that machine learning approaches have better efficiency (precision) with attribute selection approaches. We then note that all learning models outperform textual models (Lucene Solr model and BM25) as well as our first propositions based on the linear combination approach. We finally find that the J48 decision tree is the most appropriate model, it takes into consideration all the social signals,

the improvement rates compared to Naive Bayes and SVM are 13% and 15% on IMDb as well as 6% and 17% on SBS, respectively. In addition, the J48 gives the best improvements over all previous approaches, the improvement rate compared to the default model BM25 (configuration named Lucene Solr in Tables 8 and 9) is 144% for the IMDb collection and 152% for the SBS collection, while comparing to the best results obtained by the model based on linear combination (configuration named "All Properties" in Tables 8 and 9) is 27% for the IMDb collection and 63% for the SBS collection.

Finally, all these experiments clearly show that social signals allow to enhance a search. These improvements show the interest of social relevance, knowing that qualitative properties (*popularity* and *reputation*) and the temporal property (*freshness*) provide a significant improvement compared to the configuration ignoring these properties (textual model only or signals taken individually). We observe that the resources having more positive data (e.g., *like*, *+1*, *rating*) are trustworthy than the ones don't possess these social signals. If multiple users have found that the resource is useful, then it is more probable that other users will find these resources useful too. After these experiments, we observe that learning models are much more suitable than linear combination on exploiting of this type of social signals to enhance a search. We can say that the J48 learning model with selection attribute algorithm improves a precision of search results significantly.

4.6. Ranking Correlation Analysis

In order to analyze social signals and determine if there is a link (dependence / independence) between them and the document relevance, thus that between them in pairs, we conducted a correlation study. Our goals are as follows:

- first, determine the social signals that are correlated with the relevance, and facilitate the interpretation of the results.
- second, determine the redundant signals, and those that have a same effect on the retrieval improvement.

4.6.1. Correlation Between Signals and Relevance

According to a June 2014 study from Searchmetrics¹², among 22 ranking factors identified, social

signals account for 5 of the 6 most highly correlated with Google search results. In addition, BrightEdge¹³ survey released in 2013, 84% of search marketers say social signals such as *like*, *tweet*, and *mention +1* will be either more important (53%) or much more important (31%) to their SEO (Search Engine Optimization) compared to previous years.

Social signals continue to become more and more a highly correlated factor with the results of Google. Although we did not see a lot of scientific studies on these signals, some marketing organizations such as Searchmetrics continue to analyze them. According to the 2015 Searchmetrics study [47], the correlations of social signals rankings are practically unchanged compared to 2014 and remains at a high level. The first results returned by Google contain more social signals, this factor increases exponentially in the first places. Figure 4 shows the 2015 results of the correlation between social signals and Google results.

We analyzed the ranking correlation between signals and relevance using a correlation coefficient of Spearman's Rho (r_s), that measures the agreement between each signal value and document relevance class [19]. If the value of *Rho* is positive, we can say that there is some correlation between the two variables. The value of *Rho* is between $[-1, 1]$, more the *Spearman Rho* is close to 1, more the relation is strong and vice versa [19].

Figure 5 shows the values of correlations between ranges social signals (individually and grouped as properties: *popularity* and *reputation*) with respect to documents relevance. This study shows that *Facebook like* (0.29) has the highest correlation among the other individual signals, followed by number of *Facebook comment* (0.28). Other high-ranking factors include *Facebook share* (0.27) and *tweet* (0.23). Concerning the signals grouped as properties as well as the total of Facebook signals, are the most correlated with relevance compared to signals taken individually. However, the *popularity* generates the highest correlation compared to the *reputation* and total Facebook signals.

Finally, the ranking correlation analysis shows that all social signals are positively correlated with relevance. This study justifies our hypothesis and the results obtained above (see table 8) and confirms the interest of social signals exploited: Well positioned resources have a high number of *like*, *share* and specific resources stand out in the top search results with a very

¹²www.searchmetrics.com/knowledge-base/ranking-factors-infographic-2014/

¹³<http://www.marketingcharts.com/uncategorized-20695>

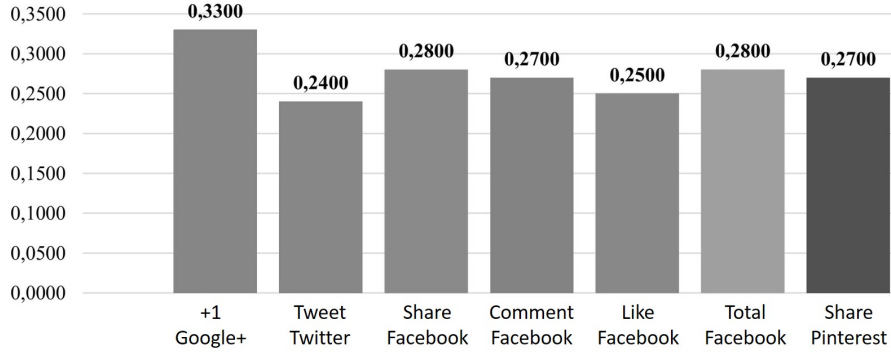


Fig. 4. Correlation between social signals and Google search results

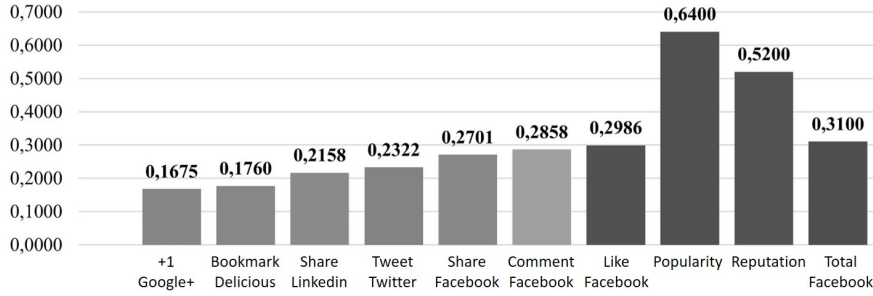
Fig. 5. *Rho* Correlations between social signals and documents relevance

Table 14
Spearman's *Rho* correlation values for the social signals pairs

Signals	Like	Share	Comment	Share(LIn)	Tweet	Bookmark	+1
Like	1						
Share	0.61	1					
Comment	0.31	0.26	1				
Share(LIn)	0.35	0.41	0.40	1			
Tweet	0.32	0.28	0.39	0.77	1		
Bookmark	0.34	0.48	0.51	0.31	0.76	1	
Google + 1	0.34	0.61	0.40	0.32	0.30	0.71	1

high mass of social data. On the one hand, this means that the activity on social networks continues to increase, on the other hand, it means that the frequently *liked* or *shared* content is increasingly correlated with good ranking of relevance.

4.6.2. Pair-wise Correlation Between Social Signals

To examine the linear relationship for each pair of social signals, we compute the pairwise overlap between the features by averaging the similarity of their top-1000 rankings over all queries. Atypical method for measuring the similarity of two ranked lists is using the Spearman's *Rho* metric. The more *Rho* is close

to 1 (in absolute value), the more the relation is strong and vice-versa.

In table 14, we provide the Spearman's *Rho* scores that are normalized to [0,1] range where 0 means completely different rankings and 1 means equal rankings. The lower diagonal of the table presents the correlation of social signals based on the rankings for all queries. We find that, the top-1000 rankings provided by the social signals pairs (*tweet*, *share(LIn)*), (*bookmark*, *tweet*) and (*mention +1*, *bookmark*) are highly correlated, i.e., the similarity scores of these pairs are higher than 0.70 (see table 14). These correlations between social signals imply some redundancy, at least

for the purposes of ranking. These observations justify and confirm the results obtained by using feature selection approach to filter and rank such redundant social signals. In this study, social signals: *bookmark*, *share* (*LIn*) are the less important criteria followed by *+1*.

Finally, this is a preliminary correlation study, we are well aware that further reflection to better address these issues is needed.

5. Conclusion

This paper proposes a search model exploiting social signals. These signals (User Generated Content), collected from several social networks, can quantify some social properties such as *popularity*, *reputation* and *freshness*. The proposed model combines linearly two relevance scores: (1) topical, estimated using classical IR model and (2) social, estimated using some social features, *popularity*, *reputation* and the *freshness* of resources. Experimental evaluation conducted on two INEX datasets IMDb and SBS shows that the integration of social signals and their properties within a textual search model allows to improve the quality of the search results. Our evaluations using attributes selection algorithms and three state-of-the-art learning algorithms support our hypothesis: the rankers based on the social signals, including both the popularity, the freshness and the reputation outperform those built by using only basic textual features. We found that J48 brings the best improvement in terms of effectiveness compared to baseline and all our other proposed configurations. Analyzing ranking correlations, we note that all social signals present a positive correlation. Meanwhile, this correlation agreement justifies the significant improvement for our proposed social approach.

For future research, we plan to address some limitations of the current study. We plan to integrate other social data into a proposed approach (emotions, event reactions, etc.). Also, we plan to study the importance of social networks and social actors of these signals and their impact on the relevance. Further experiments on other types of collections are also needed. 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. This is even with these simple elements, the first results encourage us to invest more this track.

References

- [1] F. Abel, N. Henze, D. Krause, and M. Kriesell. *On the Effect of Group Structures on Ranking Strategies in Folksonomies*, pages 275–300. Springer Berlin Heidelberg, 2009.
- [2] R. Agrawal, B. Golshan, and E. Papalexakis. Whither social networks for web search? In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, pages 1661–1670, New York, NY, USA, 2015. ACM.
- [3] K. Albishre, Y. Li, and Y. Xu. Query-based automatic training set selection for microblog retrieval. In *Advances in Knowledge Discovery and Data Mining - 22nd Pacific-Asia Conference, PAKDD 2018, Melbourne, VIC, Australia, June 3-6, 2018, Proceedings, Part II*, pages 325–336, 2018.
- [4] O. Alonso and V. Kandylas. A study on placement of social buttons in web pages. *CoRR*, abs/1410.2828, 2014.
- [5] I. Badache. RI sociale: intégration de propriétés sociales dans un modèle de recherche. In *Conférence francophone en Recherche d'Information et Applications - CORIA 2013*, pages pp. 1–6, Neuchâtel, Switzerland, April 2013.
- [6] I. Badache. *Recherche d'information sociale : exploitation des signaux sociaux pour améliorer la recherche d'information. (Social information retrieval : harnessing social signals to enhance a search)*. PhD thesis, Paul Sabatier University, Toulouse, France, 2016.
- [7] I. Badache. Exploring differences in the impact of users' traces on arabic and english facebook search. In *IEEE/WIC/ACM International Conference on Web Intelligence*, WI '19, pages 225–232, New York, NY, USA, 2019. ACM.
- [8] I. Badache. Users' traces for enhancing arabic facebook search. In *Proceedings of the 30th ACM Conference on Hypertext and Social Media*, HT '19, pages 241–245, NY, USA, 2019. ACM.
- [9] I. Badache, A. Abu-Thaier, M. Hamdan, and L. Abu-Jaish. Social Information Retrieval in Arabic Language: Case of Facebook. In *Conférence en Recherche d'Information et Applications*, Lyon, France, March 2019.
- [10] I. Badache and M. Boughanem. Exploitation de signaux sociaux pour estimer la pertinence a priori d'une ressource. In *Conférence francophone en Recherche d'Information et Applications (CORIA)*, pages pp. 163–178, Nancy, France, 2014.
- [11] I. Badache and M. Boughanem. Harnessing social signals to enhance a search. In *Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) - Volume 01*, WI-IAT '14, pages 303–309, Washington, DC, USA, 2014. IEEE Computer Society.
- [12] I. Badache and M. Boughanem. Social priors to estimate relevance of a resource. In *Proceedings of the 5th Information Interaction in Context Symposium*, IiX '14, pages 106–114, New York, NY, USA, 2014. ACM.
- [13] I. Badache and M. Boughanem. Pertinence a Priori Basée sur la Diversité et la Temporalité des Signaux Sociaux. In *Conférence francophone en Recherche d'Information et Applications (CORIA)*, pages pp. 23–38, Paris, France, 2015.
- [14] I. Badache and M. Boughanem. A priori relevance based on quality and diversity of social signals. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '15, pages 731–734, New York, NY, USA, 2015. ACM.

- [15] I. Badache and M. Boughanem. Emotional social signals for search ranking. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '17, pages 1053–1056, New York, NY, USA, 2017. ACM.
- [16] I. Badache and M. Boughanem. Les Signaux Sociaux émotionnels : Quel impact sur la recherche d'information ? In *Conférence en Recherche d'Information et Applications (CORIA 2017)*, Marseille, France, March 2017.
- [17] S. Bao, G. Xue, X. Wu, Y. Yu, B. Fei, and Z. Su. Optimizing web search using social annotations. In *Proceedings of the 16th international conference on World Wide Web*, pages 501–510. ACM, 2007.
- [18] L. Ben Jabeur. *Leveraging social relevance: Using social networks to enhance literature access and microblog search*. PhD thesis, Université de Toulouse, Université Toulouse III-Paul Sabatier, 2013.
- [19] S. D. Bolboaca and L. Jantschi. Pearson versus spearman, kendall's tau correlation analysis on Structure-Activity relationships of biologic active compounds. *Leonardo Journal of Sciences*, 5(9):179–200, 2006.
- [20] F. Bravo-Marquez, E. Frank, and B. Pfahringer. Transferring sentiment knowledge between words and tweets. *Web Intelligence*, 16(4):203–220, 2018.
- [21] L. Chan, S. Jhaver, and S. Soni. Postscholar: Surfacing social signals in google scholar search. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*, CSCW '16 Companion, pages 17–20, New York, NY, USA, 2016. ACM.
- [22] S. Chelaru, C. Orellana-Rodriguez, and I.S. Altingovde. How useful is social feedback for learning to rank youtube videos? *World Wide Web*, 17(5):997–1025, Sep 2014.
- [23] W.S. Gosset. The probable error of a mean. *Biometrika*, 6(1):1–25, 1908.
- [24] L. Gou, X. Zhang, H. Chen, J. Kim, and C.L. Giles. Social network document ranking. In *Proceedings of the 10th Annual Joint Conference on Digital Libraries*, JCDL '10, pages 313–322, New York, NY, USA, 2010. ACM.
- [25] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1-3):389–422, 2002.
- [26] M.A. Hall. *Correlation-based feature selection for machine learning*. PhD thesis, University of Waikato Hamilton, 1999.
- [27] M.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.
- [28] B. Hecht, J. Teevan, M.R. Morris, and D.J. Liebling. Search-buddies: Bringing search engines into the conversation. In *Proceedings of the Sixth International Conference on Weblogs and Social Media*, pages 138–145, 2012.
- [29] R.C. Holte. Very simple classification rules perform well on most commonly used datasets. *Machine learning*, 11(1):63–90, 1993.
- [30] L. Hong, O. Dan, and B.D. Davison. Predicting popular messages in twitter. In *Proceedings of the 20th International Conference Companion on World Wide Web*, WWW '11, pages 57–58, 2011.
- [31] A. Hotho, R. Jäschke, C. Schmitz, editor="Sure York Stumme, Gerd", and J. Domingue. Information retrieval in folksonomies: Search and ranking. In *The Semantic Web: Research and Applications*, 3rd European Semantic Web Conference, ESWC 2006, Budva, Montenegro, June 11-14, 2006, *Proceedings*, pages 411–426, 2006.
- [32] L. Huan and S. Rudy. Chi2: feature selection and discretization of numeric attributes. In *Proceedings of 7th IEEE International Conference on Tools with Artificial Intelligence*, pages 388–391, Nov 1995.
- [33] L. Huan and S. Rudy. A probabilistic approach to feature selection - A filter solution. In *Machine Learning, Proceedings of the Thirteenth International Conference (ICML '96)*, Bari, Italy, July 3-6, 1996, pages 319–327, 1996.
- [34] V. Jan, W.L. Kenneth, and N. Wilfred. Searching for quality microblog posts: Filtering and ranking based on content analysis and implicit links. In *Database Systems for Advanced Applications - 17th International Conference, DASFAA 2012, Busan, South Korea, April 15-19, 2012, Proceedings, Part I*, pages 397–413, 2012.
- [35] B. Karweg, C. Huetter, and K. Böhm. Evolving social search based on bookmarks and status messages from social networks. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, CIKM '11, pages 1825–1834, 2011.
- [36] A. Khodaei and C. Shahabi. Social-textual search and ranking. In *Proceedings of the First International Workshop on Crowdsourcing Web Search*, Lyon, France, pages 3–8, 2012.
- [37] A. Khodaei, S. Sohangir, and C. Shahabi. *Personalization of Web Search Using Social Signals*, pages 139–163. Springer International Publishing, Cham, 2015.
- [38] R. Kohavi and G.H. John. Wrappers for feature subset selection. *Artificial intelligence*, 97(1):273–324, 1997.
- [39] W. Lewoniewski, R.C. Härting, K. Węcel, C. Reichstein, and W. Abramowicz. Application of seo metrics to determine the quality of wikipedia articles and their sources. In Robertas Damaševičius and Giedrė Vasiljeviėnė, editors, *Information and Software Technologies*, pages 139–152, Cham, 2018. Springer International Publishing.
- [40] R.S. Marko and K. Igor. An adaptation of relief for attribute estimation in regression. In *Proceedings of the Fourteenth International Conference on Machine Learning (ICML 1997)*, Nashville, Tennessee, USA, July 8-12, 1997, pages 296–304, 1997.
- [41] M.R. Morris and J. Teevan. Exploring the complementary roles of social networks and search engines. In *Microsoft Research, Human-Computer Interaction Consortium Workshop(HCIC)*, pages 1–10, 2012.
- [42] P. Pantel, M. Gamon, O. Alonso, and K. Haas. Social annotations: Utility and prediction modeling. In *Proceedings of the 35th SIGIR Conference on Research and Development in Information Retrieval*, pages 285–294, 2012.
- [43] D.R. Pierce, D.P. Redlawsk, and W.W. Cohen. Social influences on online political information search and evaluation. *Political Behavior*, 39(3):651–673, Sep 2017.
- [44] J.R. Quinlan. *C4. 5: programs for machine learning*. Elsevier, 2014.
- [45] M. Raza. A new level of social search: Discovering the user's opinion before he make one. Technical report, Cambridge, UK: Mi-crosoft Research Cambridge.
- [46] S.E. Robertson and S. Walker. Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '94, pages 232–241, 1994.

- [47] M. Tober, D. Furch, K. Londenber, L. Massaron, and J. Grundmann. Search ranking factors and rank correlations. *SearchMetrics*, 2015.
- [48] S. Yen and Y. Lee. Under-sampling approaches for improving prediction of the minority class in an imbalanced dataset. *Intelligent Control and Automation*, pages 731–740, 2006.
- [49] Q. Yuan, G. Cong, and N.M. Thalmann. Enhancing naive bayes with various smoothing methods for short text classification. In *Proceedings of the 21st International Conference on World Wide Web, WWW '12 Companion*, pages 645–646, New York, NY, USA, 2012. ACM.
- [50] Y. Yusuke, A. Jatowt, N. Satoshi, and T. Katsumi. Can social bookmarking enhance search in the web? In *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL '07*, pages 107–116, New York, NY, USA, 2007. ACM.
- [51] S. Zhang and N. Cabage. Does SEO matter? increasing classroom blog visibility through search engine optimization. In *46th Hawaii International Conference on System Sciences, HICSS 2013, Wailea, HI, USA, January 7-10, 2013*, pages 1610–1619, 2013.