Harnessing Social Signals to Enhance a Search

Ismail Badache, Mohand Boughanem

Harnessing Social Signals to Enhance a Search

Ismail Badache
IRIT UMR 5505 CNRS, France
University of Toulouse
Email: Ismail.Badache@irit.fr

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

***Abstract*—This paper describes an approach of information retrieval which takes into account social signals associated with Web resources to estimate its relevance to a query. We show how these data, which are in the form of actions within social activities (e.g. like, tweet), can be exploited to quantify social properties such as popularity and reputation. We propose a model that combines the social relevance, estimated from these properties, with the conventional textual relevance. We evaluated the effectiveness of our approach on IMDb dataset containing 32706 resources and their social characteristics collected from several social networks. We used also the selected criteria to learn models to determine their effectiveness in information retrieval. Our experimental results are promising and show the interest of integrating social signals in retrieval model to enhance a search.**

Keywords*—*Social Signals, Social Properties, Social Information Retrieval, Criteria Evaluation, Learning Models.

I. Introduction

Nowadays, social Web has completely changed the manner in which people communicate and share information in the Web. It allows users to interact and produce a large masses of social signals. In 2013, Facebook marked more than 1130 billion *like*[1], knowing that more than 2,500,000 Websites use *like* button. In 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 games[2]. Other types of functions such as *endorsement, share, comment* and *rating* allow users to interact with Web resources. Through these social actions, some resources 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. Typically, we focus, in our case, on exploiting social signals in order to estimate the resource relevance to query. The research questions addressed in this paper are the following:

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 social 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 a model search?

The remainder of this paper is organized as follows. First, we present some related work regarding social search and using social networks in IR. Then, we describes our social approach. After, in experimental section, we evaluate the effectiveness of our proposed approach and discuss the results. Finally, we conclude the paper and announce some future work.

II. Related Work

In this section, we report related work exploiting social signals to measure a priori importance of resource.

Some approaches focus on how to improve information retrieval (IR) effectiveness by exploiting users' actions and their underlying social network. *Chelaru et al.* [4] study the impact of social signals (*like, dislike, comment*) on the effectiveness of search on YouTube. They show 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 algorithm and learning to rank algorithms. *Karweg et al.* [12] propose 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, *Khodaei and Shahabi* [13] propose a ranking approach based on several social 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.* [11] use *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.

Finally, there are other studies initiated by Microsoft Bing researchers [14], [16] that show the usefulness of different social contents generated by the network of user's friends on Facebook. *Pantel et al.* [15] study the leverage of social annotation on the quality of search results. They observe 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.* [10] present 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.* [8] propose 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, authors suggest that a user can enhance search by joining a larger social networks, having more friends, and connecting larger communities.

Our goal aims to exploit social signals to improve accuracy and relevance of convention 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 in the performance.

III. Social Relevance of a Resource

Our IR approach consists of exploiting social signals to define social properties to take into account in retrieval model. We associate to each Web resource a priori relevance based on these social properties. This relevance is then combined with a classical topical relevance (see Figure 1).

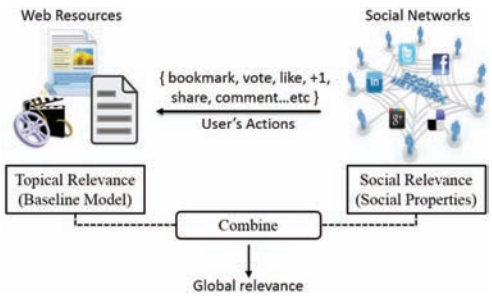

Fig. 1. A modular approach for Social IR

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

Actions. There is a set $A=\{a_1, a_2,...a_m\}$ of m actions that users can perform on the resources. These actions represent the relation between users $U=\{u_1, u_2,...u_l\}$ 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 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 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 $G=\{g_1, g_2,...g_z\}$ of z social networks. Each specific social network contains specific social actions a performed on resources.

B. Formalization of Search Model

By analyzing various types of social actions (or data) 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:

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

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

Where:

- $Count(a_i^x, r, G)$ represents number of occurrence of specific action a_i^x performed on a resource in relation to a specific social network. $x = \{P, R\}_{Social}$.

- $f_x(r, G)$ arithmetic function that represents the linear combination of m social actions that quantify each x social properties (rate of interaction through the social signals).
- $Min(f_x(r, G))$, $Max(f_x(r, G))$ represent the minimum and maximum value for f_x. $f_x(r, G)_{Norm}$ 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 publication 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 2013, the same resource was discussed in 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 initialised 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:

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

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

IV. Experimental Evaluation

To evaluate our model, we conducted a series of experiments on IMDb dataset. We first evaluated the impact of social signals, taken separately and when they are combined as the reputation and popularity. Secondly, we study the effectiveness of each social signal using machine learning with selection attributes algorithms. We show in what follows our experimental protocol.

A. Description of Test Dataset

We collected 32706 English documents extracted from "imdb.com". Each document describes a movie, and is represented by a set of metadata, and has been indexed according to keywords extracted from fields with status $indexed$ in Table I. For each document, we collected specifics social signals via their corresponding API of 5 social networks listed in Table II. We have put them in the UGC (User Generated Content) field. This field has not been indexed. The nature of these social signals is a counting of each specific social action on a resource.

TABLE I. List of the different document metadata field

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

TABLE II. Social signals statistics in the dataset

Social Networks	Social signals	Sum	Min	Max	Average
Facebook	Like	5056517	0	79693	154
	Share	5778414	0	41618	176
	Comment	6717573	0	60081	205
Twitter	Tweet	1097204	0	22954	33
Google+	Mention +1	139189	0	1368	4
Delicious	Bookmark	32810	0	1033	1
LinkedIn	Share	57545	0	25215	1

We note that we have not been able to use INEX IMDb dataset because the documents do not contain social signals.

B. Topics and Relevance Judgments

We chose 30 topics from the set of INEX IMDb topics (see Table III). To obtain relevance judgments, we use Qrels provided by INEX IMDb. The test dataset contains some documents not judged in INEX IMDb Qrels. Indeed, we noticed that for 300 documents returned by query, there is on average 33 documents not judged. For these, we asked 12 users to assess their relevance manually. Among assessors, 2 are assistant professors, 3 are PhD students, 4 are Master students and 3 are engineers. All participants were from computer science related disciplines.

To evaluate the documents displayed for a given query, we use 3-point rating scale, i.e., in the order of irrelevant, little

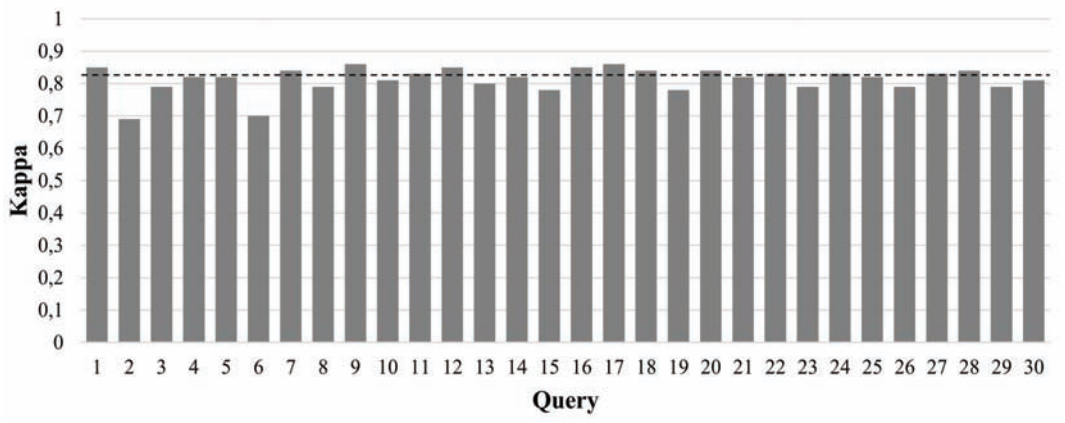

Fig. 2. Distribution of the Kappa measure k per query. < 0 poor agreement, 0.0 - 0.2 slight agreement, 0.21 - 0.4 fair agreement, 0.41 - 0.6 moderate agreement, 0.61 - 0.8 substantial agreement, 0.81 - 1 perfect agreement.

TABLE III. INSTANCES OF INEX IMDB TEST TOPICS

Topic	Description
action biker	Search for all action movies with bikers in it.
true story drugs +addiction -dealer	Find movies about drugs (drug addiction but not drug dealers) that are based on a true story.
ancient Rome era	find the movies about the era of ancient Rome.

relevant and relevant. We note that each topic is judged by 3 users. To avoid any bias, no social signals were displayed along with the documents, but all the textual metadata are displayed to facilitate the judgment task.

We have analysed the agreement degree between assessors for each topic with the Kappa measure k [5]. This indicator takes into account the proportion of agreement between assessors $\bar{P}$ and the proportion of expected agreement between assessors by chance $\bar{P}_e$. The Kappa measure is equal to 1 if assessors always agree, 0 if they agree only by chance. k is negative if the agreement between assessors is worse than random. k measure is computed as follows:

$$k = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e}$$
$$with : \bar{P} = \frac{1}{n}\sum_{i=1}^{r} n_{ii} \;\; and \;\; \bar{P}_e = \frac{1}{n^2}\sum_{i=1}^{r} n_{i1}n_{i2} \quad (5)$$

Let n be the total number of assessments supplied by the whole assessors, r be the number of assessment categories (in our case 3 categories: 0, 1 and 2). n_{ii} is the number of agreements between the two assessors for agreement i with $i \in r$, n_{i1} is the total of assessments i given by assessor 1 and n_{i2} is the total of assessments i given by assessor 2. Figure 2 shows the distribution of the kappa measure according to the query set. We notice the agreement measure ranges from 0.69 to 0.86. The average agreement measure between assessors is 81.23%, which corresponds to good agreement.

C. Quantifying Social Properties

Table IV presents the properties that we want to take into account in our retrieval model. In order to quantify these social properties, we associate them the corresponding social signals.

TABLE IV. EXPLOITED SOCIAL SIGNALS LIST IN THE QUANTIFICATION

Social properties	c_i	Social signals exploited	Social Networks
Popularity	c_1	Number of *Comment*	Facebook
	c_2	Number of *Tweet*	Twitter
	c_3	Number of *Share(LIn)*	LinkedIn
	c_4	Number of *Share*	Facebook
Reputation	c_5	Number of *Like*	Facebook
	c_6	Number of *Mention* +1	Google+
	c_7	Number of *Bookmark*	Delicious
Freshness	c_8	*Date of last action*	Facebook

Specific social signals (actions) have been associated with each property depending on their nature and meaning. In Table IV, 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 *Mention* +1, 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 date of the different actions are not available except the last date of Facebook actions (*comment* and *share*).

D. 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 (6) is the weighted sum of social and topical relevance:

$$Rel(q, r, G) = \alpha \cdot Rel_T(q, r) + (1 - \alpha) \cdot Rel_S(q, r, G) \quad (6)$$

Where: $\alpha \in$ [0, 1] as a weighting parameter and $Rel_T(q, r)$ is the normalized score of topical relevance. We used BM25 [17]

and Lucene Solr scoring[3] 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 6) 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.

TABLE V. COMPARING RETRIEVAL EFFECTIVENESS TO LUCENE SOLR DEFAULT MODEL AND BM25 MODEL

	P@10	P@20	nDCG@10	nDCG@20
BM25	0.2912	0.2276	0.3158	0.3466
Lucene Solr	0.2617	0.1951	0.2744	0.3180
Like	0.3963**	0.3142**	0.4678**	0.4873**
Share	0.4037**	0.3119**	0.4857**	0.4911**
Comment	0.3855**	0.3004**	0.4282**	0.4103**
Tweet	0.3488**	0.2613**	0.3765**	0.4096**
Mention+1	0.3206**	0.2467**	0.3478**	0.3627**
Share(LIn)	0.3177*	0.2411*	0.3461*	0.3509*
Bookmark	0.3358*	0.2395*	0.3619*	0.3517*
All Criteria	0.4225**	0.3702**	0.5111**	0.4974**
Freshness	0.3729**	0.2901**	0.4243**	0.4575**
Reputation	0.4167**	0.3622**	0.5225**	0.5488**
Popularity	0.4513**	0.3961**	0.5343**	0.5478**
Reputation+Freshness	0.4379**	0.3788**	0.5300**	0.5481**
Popularity+Freshness	0.4605**	0.4023**	0.5417**	0.5604**
Popularity+Reputation	0.4835**	0.4267**	0.5541**	0.5766**
All Properties	**0.4984****	**0.4372****	**0.5623****	**0.5971****

Table V summarizes the results of precision and nDCG [6] for 10 and 20 top documents. We evaluated different configurations, by taking into account social signals individually and their combination in the form of social properties. In order to check the significance of the results, we performed the Student test [7] and attached * (strong significance against BM25) and ** (very strong significance against BM25) to the performance number of each row in the table V when the p-value < 0.05 and p-value < 0.01 confidence level, respectively.

We observe in all cases, with taking into account social features, the results are significantly better compared to textual models. It is clear that combining social signals as social properties provides better results than when they are taken individually. The results show that *popularity* oriented signals provides better results than *reputation* oriented signals. The freshness in our study is seen in relation to the recency of actions in social activities, the number of actions on a resource is related to its freshness in social networks, but the resources that possess fresher signals are assumed to be better classified. 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 more effective to improve precision and nDCG. Therefore, combination of *freshness* with *popularity* and *reputation* provides the best improvement compared to a random combination of all criteria. Finally, we also found that taking into account the *freshness* only and combined with other properties also improves the results. It is also noted that the *freshness* in our case is correlated with the presence of social signals. So one question remains, is it just the presence of the signal that improves the relevance or it is the *freshness* of the signal that also contributed to this improvement. More detailed analysis are needed to answer clearly this question.

We can explain these results by the positive sense of *reputation* property quantified through the counting of *like*, *mention+1* and *bookmark*, 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.

E. Result of Machine Learning Study

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 IV. 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.

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[9] aim to identify and eliminate as many irrelevant and redundant information as possible. We used Weka[4] 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 300 resources for 30 topics were extracted using Lucene Solr model. Then, the scores of all criteria (social signals) are calculated for each resource. We identify relevant resources and irrelevant according to the qrels. The set of resources obtained contains 9000 instances composed of 2068 relevant resources and 6932 irrelevant resources. 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 [18]. 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 2068 relevant resources and 2068 irrelevant resources that were randomly selected. Finally, we applied the attribute selection algorithm on the whole set.

[3] http://lucene.apache.org/solr/

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

TABLE VI. SELECTED SOCIAL SIGNALS WITH ATTRIBUTE SELECTION ALGORITHMS

Algorithm	Metric	LS	c_1^{++}	c_2^{+}	c_3	c_4^{++}	c_5^{++}	c_6^{+}	c_7	c_8
CfsSubsetEval	[folds Number]	5	5	5	-	5	5	2	-	-
WrapperSubsetEval	[folds Number]	5	1	1	1	4	5	1	3	1
ConsistencySubsetEval	[folds Number]	5	5	5	5	5	5	5	4	-
FilteredSubsetEval	[folds Number]	5	5	5	-	5	5	2	-	-
ChiSquaredAttributeEval	[Rank]	1	4	5	7	2	3	6	8	9
FilteredAttributeEval	[Rank]	1	4	5	7	2	3	6	8	9
GainRatioAttributeEval	[Rank]	1	2	5	8	3	4	6	7	9
InfoGainAttributeEval	[Rank]	1	4	5	7	2	3	6	8	9
OneRAttributeEval	[Rank]	1	3	5	7	4	2	6	8	9
ReliefFAttributeEval	[Rank]	1	4	8	6	2	3	5	7	9
SVMAttributeEval	[Rank]	1	6	7	3	2	5	4	8	9
SymetricalUncertEval	[Rank]	1	2	5	7	3	4	6	8	9
	Count	12	12	12	10	12	12	12	10	9

Table VI shows the social signals selected through selection attribute algorithms. We use ranking methods to rank the selected criteria. The line "number of folds" in the table indicates how many times the criterion has been selected in the cross-validation task.

To evaluate the learned models, we used Lucene Solr results of all topics as training set and we apply a Cross-validation for 5-folds. Learned model predicts the relevance class (relevant or irrelevant) for resulting resources and give effectiveness classification scores. The irrelevant predicted resources are then deleted and resources predicted as relevant are ranked given the effectiveness classification scores (probability distribution of classes predictions). To evaluate our runs, we used the P@20 measure.

We chose to experiment with three learning algorithms. Authors in [9] studied the effectiveness of some selection attribute approaches with learning algorithms. Based on their study, we used the same combination of learning and attribute selection approaches applied on our own criteria (see Table VII) [19]:

- Naive Bayes and WrapperSubsetEval (WRP). All criteria were selected.
- Naive Bayes and CfsSubsetEval (CFS). The criteria selected in this case were: Lucene Solr score, c_1, c_2, c_4, c_5, c_6.
- J48 and ReliefFAttributeEval (RLF). In this case all criteria were selected.
- SVM and SVMAttributeEval (SVM) that assess attributes using the SVM classifier. In this case all criteria were selected.

TABLE VII. MACHINE LEARNING RESULTS (P@20)

Classifiers	Attribute selection criteria	All criteria
NaiveBayes	**0.5315** (CFS) 0.5105 (WRP)	0.5105
SVM	0.5131 (SVM)	0.5131
J48	**0.6890** (RLF)	**0.6890**

Our aim for this study is to check if attribute selection improves effectiveness of learned models. We observe that only CFS algorithm confirms clearly the hypothesis. We could see that machine learning approaches have better effectiveness with attribute selection approaches, except that all the criteria are selected by the algorithm (SVM, RLF and WRP). We notice also that all learning models outperforms textual models (Lucene Solr model and BM25) and our first linear combination approaches. J48 is the most suitable model and gives the best improvement. We also note that the criteria c_3*: share(LIn)*, c_7*: bookmark* and c_8*: freshness* are less favoured by selection algorithms. The criteria c_6: +1 and c_2: *tweet* are moderately favoured (+) but in each iteration are selected which indicates their importance even if they are not the best. Thus, the facebook criteria were the highest ranked (++) and often validated over the 5 iterations of cross validation.

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*) provide more gain compared to temporal property. We observe that the resources having more positive data (*like, mention+1*) 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.

F. Ranking Correlation Analysis

According to a June 2013 study from Searchmetrics[5], social signals account for 5 of the 6 most highly correlated with Google search results. In addition, BrightEdge[6] survey released in 2012, 84% of search marketers say social signals such as *like*, *tweet*, and +1 will be either more important (53%) or much more important (31%) to their SEO (Search Engine Optimization) compared to 2011.

We analyzed the ranking correlation performed through the Spearman's Rho (r_s) rank correlation coefficient [3], that measure the agreement between each social signals and documents relevance.

Figure 3 shows the values of correlations between ranges social signals with respect to documents relevance. The study shows that *Facebook share* (0.29) has the highest correlation, followed by number of *comment* (0.28). Other high-ranking factors include *like* (0.27) and *tweet* (0.23).

[5]www.searchmetrics.com/en/services/ranking-factors-2013/

[6]www.marketingcharts.com/direct/social-signals-increasingly/

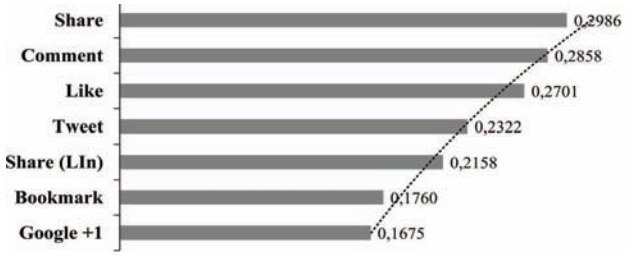

Fig. 3. Spearman correlation between social signals and relevance

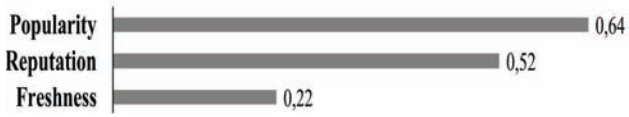

Fig. 4. Spearman correlation between social properties and relevance

In Figure 4, we also measured the correlation between ranges social properties (*popularity*, *reputation* and *freshness*) with respect to documents relevance.

The ranking correlation analysis shows that all social signals are positively correlated. Our study confirms the interest of social signals exploitation: well positioned resources have a high number of *like*, *share* and specific resources stand out in the top search results with a very high mass social data. On 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.

V. Conclusion

This paper proposes a search model exploiting social properties. These properties have been defined on the basis of social signals (User Generated Content) collected from several social networks. The proposed model combines linearly two relevance scores: (1) topical, estimated by a classical IR model; (2) social, estimated by the *popularity*, *reputation* and the *freshness* of resources. Experimental evaluation conducted on the IMDb dataset shows that the integration of social properties within a textual search model allows to improve the quality of the search results. We then attempted to compare some well-known learning approaches and we found that J48 brings the best improvement in terms of effectiveness compared to baseline and our proposed approaches. Analyzing ranking correlations, we note that all social signals present a positive correlation. Meanwhile, this correlation agreement justifies the significant improvement for our social models.

For future research, we plan to address some limitations of the current study. We plan to integrate other social data into a proposed approach. Further experimentations on other types of collections are also needed. This is even with these simple elements, the first results encourage us to invest more this track.

References

[1] I. Badache, "RI sociale: intgration de proprits sociales dans un modle de recherche," in Proceeding of 10th French Information Retrieval Conference, ser. CORIA'13. Neuchâtel, Switzerland: ARIA, 2013, pp. 463-468.

[2] I. Badache, M. Boughanem, "Exploitation de signaux sociaux pour estimer la pertinence a priori d'une ressource," in Proceeding of 11th French Information Retrieval Conference, ser. CORIA'14. Nancy, France: ARIA, 2014, pp. 163-178.

[3] S.D. Bolboaca, L. Jntschi, "Pearson versus Spearman, Kendall's tau correlation analysis on structure-activity relationships of biologic active compounds," Leonardo Journal of Sciences, vol. 5, n. 9, pp. 179-200, 2006.

[4] S. Chelaru, C. Orellana, I. Altingovde, "Can Social Features Help Learning to Rank Youtube Videos?," in Proceedings of the 13th International Conference on Web Information Systems Engineering, ser. WISE'12. Paphos, Cyprus: Springer-Verlag, 2012, pp. 552-566.

[5] J. Cohen, "A coefcient of agreement for nominal scales ," in Educational and Psychological Measurement, vol. 20, n. 1, pp. 213-220, 1960.

[6] K. Järvelin, J. Kekäläinen, "Cumulated gain-based evaluation of information retrieval techniques," in ACM Transactions on Information Systems, vol. 20, n. 4, pp. 422-446, 2002.

[7] W. S. Gosset, "The probable error of a mean," Biometrika, vol. 6, n. 1, pp. 1-25, Originally published under the pseudonym "Student", 1908.

[8] L. Gou, X. Zhang, H. Chen, J. Kim, C. Giles, "Social Network Document Ranking," in Proceedings of the 10th Annual Joint Conference on Digital Libraries, ser. JCDL'10. Gold Coast, Queensland, Australia: ACM, 2010, pp. 313-322.

[9] M.A. Hall, G. Holmes, "Benchmarking attribute selection techniques for discrete class data mining," in Knowledge and Data Engineering, IEEE Transactions on, vol. 15, n. 6, pp. 1437-1447, 2003.

[10] B. Hecht, J. Teevan, M.R. Morris, D. Liebling, "SearchBuddies: Bringing Search Engines into the Conversation," in Proceedings of International Conference on Weblogs and Social Media, ser. ICWSM'12. USA: AAAI, 2012, pp. 138-145.

[11] L. Hong, O. Dan, B. Davison, "Predicting Popular Messages in Twitter," in Proceedings of the 20th International Conference Companion on World Wide Web, ser. WWW'11. Hyderabad, India: ACM, 2011, pp. 57-58.

[12] B. Karweg, C. Hütter, K. Böhm, "Evolving Social Search Based on Boukmarks and Status Messages from Social Networks," in Proceedings of the 20th ACM International Conference on Information and Knowledge Management, ser. CIKM'11. Glasgow, Scotland, UK: ACM, 2011, pp. 1825-1834.

[13] A. Khodaei, C. Shahabi, "Social-Textual Search and Ranking," in Proceedings of the First International Workshop on Crowdsourcing Web Search, ser. CrowdSearch'12. Lyon, France: CEUR-WS.org, 2012, pp. 3-8 .

[14] MR. Morris, J. Teevan, "Exploring the Complementary Roles of Social Networks and Search Engines," in Microsoft Research, USA, 2013, pp. 1-10.

[15] P. Pantel, M. Gamon, O. Alonso, K. Haas, "Social Annotations: Utility and Prediction Modeling," in Proceedings of Special Interest Group on Information Retrieval, ser. SIGIR'12. USA: ACM, 2012, pp. 285-294.

[16] M. Raza, "A new Level of Social Search: Discovering the User's Opinion Before he Make One," in Microsoft Research Cambridge, UK, 2011, pp. 1-6.

[17] E. Robertson, 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, ser. SIGIR'94. Dublin, Ireland: Springer-Verlag, 1994, pp. 232-241.

[18] S.-J Yen, Y.-S Lee, "Under-sampling approaches for improving prediction of the minority class in an imbalanced dataset," in International Conference on Intelligent Computing, ser. ICIC'06. China: Springer Berlin Heidelberg, 2006, vol. 344, pp. 731-740.

[19] Q. Yuan, G. Cong, N.M. Thalmann, "Enhancing naive bayes with various smoothing methods for short text classification," in Proceedings of the 21st International Conference Companion on World Wide Web, ser. WWW'12 Companion. Lyon, France, 2012, pp. 645-646.