Stories Around You A Two-Stage Personalized News Recommendation

Conference Paper · October 2014

DOI: 10.5220/0005159804730479

CITATION
1

READS
342

4 authors:

Youssef Meguebli
France Télécom
17 PUBLICATIONS **40** CITATIONS
SEE PROFILE

Mouna Kacimi
Free University of Bozen-Bolzano
45 PUBLICATIONS **492** CITATIONS
SEE PROFILE

Bich-Liên Doan
CentraleSupelec
88 PUBLICATIONS **239** CITATIONS
SEE PROFILE

Fabrice Popineau
Laboratoire de Recherche en Informatique
46 PUBLICATIONS **117** CITATIONS
SEE PROFILE

Stories Around You
A Two-Stage Personalized News Recommendation

Youssef Meguebli[1], Mouna Kacimi[2], Bich-liên Doan[1] and Fabrice Popineau[1]
[1]*SUPELEC Systems Sciences (E3S), Gif sur Yvette, France*
[2]*Free University of Bozen-Bolzano, Bolzano, Italy*

Keywords: News recommendation, Opinion mining, Diversification.

Abstract: With the tremendous growth of published news articles, a key issue is how to help users find diverse and interesting news stories. To this end, it is crucial to understand and build accurate profiles for both users and news articles. In this paper, we define a user profile based on (1) the set of entities she/he talked about it in her/his comments and (2) the set of key-concepts related to those entities on which the user has expressed an opinion or a viewpoint. The same information is extracted from the content of each news article to create its profile. In a first step, we matched those profiles using a new similarity measure. We use also the news articles profiles to diversify the list of recommended stories in a second step. A first evaluation involving the activities of 150 real users in four news web sites, namely The Independent, The Telegraph, CNN and Aljazeera has shown the effectiveness of our approach compared to recent works.

1 INTRODUCTION

News web sites like CNN[1] and Aljazeera[2] are becoming one of the main platforms where users can comment on the latest breaking news to express their opinions. Such comments contain rich information about users' convictions and interests and thus represent a valuable source for identifying users' profile, which is the key for an effective recommendation system.

The accuracy of personalized recommendation depends mainly on how well user profiles are defined. Naturally, users' comments represent a valuable information source since they reflect not only interesting topics for users but also more details about their viewpoints regarding specific issues.

Therefore, several past studies have exploited, in different ways, users' comments for news recommendation (Abbar et al., 2013; Abel et al., 2011; Weng et al., 2010; Shmueli et al., 2012; Meguebli et al., 2014a). Most of them use tweets (Abel et al., 2011; Weng et al., 2010) and few others (Abbar et al., 2013; Shmueli et al., 2012; Li et al., 2010; Meguebli et al., 2014a) exploit users' comments on news websites. For example, Li et. al., (Li et al., 2010) enrich the content of each news article using users' comments for a more enhanced recommendation. However they do not build any user profile which results in a limited accuracy. Shmueli et. al., (Shmueli et al., 2012) restrict user profile to a set of tags extracted from related comments.

The closest work to ours is by Abbar et. al.,(Abbar et al., 2013) who build the profile of each user by extracting the set of entities she/he has commented on. While the proposed approach is interesting, it does not exploit all available information in users' comments and thus it provides incomplete profiles. The reason is that a user can have an interest only on some key-concepts related to a given entity and be not interested to the same entity when it is related to other key-concepts. For example, a user can be interested to the entity Tunisia when it is related to the key-concept Tourism, while he can show no interest in the same entity when it is about the key-concept Election.

In this paper, we define a fine grained user profile described by a set of tuples (e_i, c_{ij}) representing the entities and key-concepts of interests related to each user. The entities and their related key-concepts are extracted from the comments of each user. Similarly, the profile of news article is described by a set of tuples (e_i, c_{ij}). For news articles profiles, we define also the sentiment orientation toward each tuple which will be used later to diversify the list of recommended news articles. The sentiment orientation of a

[1]http://www.cnn.com
[2]http://www.aljazeera.com

Meguebli Y., Kacimi M., Doan B. and Popineau F..
Stories Around You - A Two-Stage Personalized News Recommendation.
DOI: 10.5220/0005159804730479
In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval* (KDIR-2014), pages 473-479
ISBN: 978-989-758-048-2

given tuple can be positive, negative or neutral.

To ensure the recommendation of relevant and diverse news stories, we propose a model based on two main stages: (i) First, we select news articles that are closest to the user profile using a new similarity measure between user and news articles profiles.
(ii) Second, we diversify the list of selected news articles by applying a news articles diversification model based on two main components: (1) semantic diversification on the list of relevant news articles to avoid redundancy and cover a diverse set of news articles presenting different arguments, and (2) sentiment diversification to cover different types of sentiments that can be positive, negative or neutral.

We evaluate our approach using four real datasets including, The Independent, The Telegraph, CNN, and Aljazeera. The first experiments show that our approach outperforms a recent baseline approach with a large margin, in terms of precision and NDCG.

The rest of the paper is organized as follows. Section 2 reviews related work and puts it into context. Section 3 introduces the user and articles profiles. Section 4 shows the proposed model for news recommendation. Section 5 presents experiments that demonstrate the effectiveness of our model. Finally, section 6 concludes the paper.

2 RELATED WORK

Several approaches have been focused on user generated content to improve the effectiveness of retrieval and recommender systems (Stoyanovich et al., 2008; Abel et al., 2011; Hong and Davison, 2010; Weng et al., 2010; Michelson and Macskassy, 2010). Stoyanovich et. al., (Stoyanovich et al., 2008) leverage the tagging behavior of users to derive implicit social ties which were shown to serve as good indicator of user's interests. Chen et. al., (Chen et al., 2010) exploits user Tweets to build a bag-of-words profile for each Twitter user. Abel et al., (Abel et al., 2011) build hashtag-based, entity-based, and topic-based user profiles from Tweets, and show that semantic enrichment improves the variety and the quality of profiles. Other approaches (Hong and Davison, 2010; Weng et al., 2010; Michelson and Macskassy, 2010) address the problem of extracting topics of interest in micro-blogging environments. Hong et.al., (Hong and Davison, 2010) train a topic model on aggregated messages to improve the quality of topic detection in Tweets. Similarly, Weng et. al., (Weng et al., 2010) apply Latent Dirichlet Allocation (LDA) model to identify latent topic information from Tweets. Michelson et. al., (Michelson and Macskassy, 2010) use a knowledge base to disambiguate and categorize the entities in user Tweets and then develop users profiles based on frequent entity categories.

Our work does not fall in the previous class since we exploit richer and longer comments than tweets. Thus, we relate our work to a second class of approaches (Abbar et al., 2013; Shmueli et al., 2012; Li et al., 2010; Meguebli et al., 2014b) which exploit users' comments on news websites to build user profiles. Meguebli et al, (Meguebli et al., 2014b) identify political orientation of users based on their sentiments towards aspects extracted from their comments on news sites. Li et. al., (Li et al., 2010) enrich the content of each news article using users' comments and use the enhanced content to improve the accuracy of recommendation. However they do not build any user profile which results in a limited accuracy. Shmueli et. al., (Shmueli et al., 2012) restrict user profile to a set of tags extracted from related comments using a bag-of-words model. However, they do not take into account any sentiment information. The closest work to ours is by Abbar et. el., (Abbar et al., 2013) who build the profile of each user by extracting the set of entities she/he has commented on and their related sentiments. While the proposed approach is interesting, it does not exploit the different aspects of entities to have a more precise profile. In our work, we model the user profile as a set of viewpoints reflected by *entities* and *key-concepts*.

Another research area related to our work is search result diversification which has been investigated extensively following two different approaches. The first one is *taxonomy-independent* where no knowledge base is used to diversify search results (Zhai and Lafferty, 2006; Radlinski and Dumais, 2006; Santos et al., 2010; Gollapudi and Sharma, 2009; Wang and Zhu, 2009). Some of the works falling into this category include the work by Gollapudi et.al., (Gollapudi and Sharma, 2009) that uses a diversification model combining both novelty and relevance of search results. Radilinski et.al. (Radlinski and Dumais, 2006) use query expansion to enrich search results generating more relevant documents for various interpretations. The second approach to result diversification is *taxonomy-based*. (Agrawal et al., 2009; Clarke et al., 2008; Carterette and Chandar, 2009). Representative works include the one by Agrawal et.al., (Agrawal et al., 2009) which makes use of a taxonomy for classifying queries and documents and create a diverse set of results according to this taxonomy. Clarke et.al., (Clarke et al., 2008) focus on developing a framework of evaluation that takes into account both novelty and diversity. Carterette et.al., (Carterette and

Chandar, 2009) propose a probabilistic approach to maximize the coverage of the retrieved documents with respect to the aspects of a query. In our work, we adopt the technique proposed by Kacimi et. al., (Kacimi and Gamper, 2011), a *taxonomy-independent* approach specific for news articles diversification.

For completeness, we provide a brief survey of existing recommendation system approaches (Abbar et al., 2013; Abel et al., 2011; Shmueli et al., 2012; Li et al., 2010; Chen et al., 2010; Phelan et al., 2009) where two main strategies have been adopted. First, content filtering strategies which create a profile for each user or seed article and then recommends the best matching articles based on user profile, seed article, or both. Second, collaborative filtering strategies that rely only on past user behavior without requiring the creation of explicit profiles. In our work, we adopt a content filtering strategy to recommend news articles to users based on their profiles. Moreover, most of recommendation system techniques deal with media and entertainment products while in our work we focus on recommending news articles which are textual items subject to high volatility and churn rate.

3 MODELING USER AND ARTICLE PROFILES

3.1 User Profile

To define the profile of a given user u, we collect the opinions he has expressed, in all news sites, during a period of time T. Then, we analyze the opinions and extract from them a set of tuples $\{\{(e_1,c_{11}),(e_1,c_{12})...,(e_n,c_{nm})\}$ where e_i is an entity (e.g., Person, Location, Organization) and c_{ij} is the key-concept related to each entity e_i. The key-concept c_{ij} can be an entity attributes or some abstract objects. In this work we extract the key-concepts using ODP taxonomy[3]. For instance, we extract from the opinion *"Obama is wrong to give work permits to young illegal immigrants"* the entity Obama (Person) and their related key concepts Work permit and illegal immigration. Practically, to build a user profile, we first identify all opinions expressed by the user u for a period of time T. We identify all sentences of its content using OpenNLP[4]. Thus, for each sentence, we extract the different entities and their related key-concepts. Formally, the profile of a user u is defined by:

$$P(u) = \{(e_i,c_{ij}),w_u(e_i,c_{ij})|e_i \in E, c_{ij} \in C, u \in U\} \quad (1)$$

Where C, E and U denote the set of entities, key-concepts and users respectively and $w_u(e_i,c_{ij})$ is the weight of each tuple (e_i,cij) computed using *tf*idf* technique . In our work, *tf* represents the tuple frequency in the set of comments of the user U, and *idf* represents the inverted document frequency in the set of comments of all users.

3.2 News Article Profile

Similarly to user profile, we represent each news article by a set of tuples $< e_i,c_{ij} >$ extracted from its content. However, we define also the sentiment towards each tuple. Practically, to build a news article profile, we first identify all sentences of its content using OpenNLP. For each sentence, we define its sentiment orientation using AlchemyApi[5]. The sentiment orientation of a sentence can be positive, negative or neutral. Thus, for each news article we obtain three group of sentences corresponding to positive, negative and neutral profiles of the news article. For each group of sentences, we extract their tuples corresponding to entities and their related key-concepts. The weight of each tuple is defined through *tf*idf* technique where *tf* is the tuple frequency in the sentences of a given news article and *idf* is the inverted document frequency in all sentences of all news articles.

4 NEWS RECOMMENDATION MODEL

We propose a two stage recommendation model: In a first step, we select the topk[6] relevant news articles by computing cosine similarity between user profile and news articles profiles, where the unit item is a tuple (e_i,c_{ij}). This measure has been shown to be very effective in measuring similarity between documents (Singhal, 2001). In a standard search problem, a document is represented by a vector of n dimensions where a term is assigned to each dimension and the value of the dimension represents the frequency of the term in the document. In our setting we are interested in computing similarity between tuples, so each profile is represented by a vector where the dimensions of each vector are assigned tuples and the value of each dimension represents the *tf*idf* score of the tuple for the given profile. Formally the cosine

[3]www.dmoz.org

[4]http://opennlp.sourceforge.net/

[5]www.alchemyapi.com

[6]In this work we have set k=200

similarity between a news article profile A and a user profile B is given by:

$$Similarity(A,B) = \frac{\frac{B.A^+}{||B||||A^+||} + \frac{B.A^-}{||B||||A^-||} + \frac{B.A^o}{||B||||A^o||}}{3}$$

where B is the vector corresponding to the user profile B, and A^+, A^-, and A^o are respectively the positive, negative, and neutral vectors corresponding to the news article profile A. We compute the cosine similarity between each type of vector and then we average the results to obtain the final similarity values. The more tuples an article profile and a user profile have in common, the more interesting is the article for the user. In the second stage, we perform diversification of news articles. The technique used to diversify news articles was inspired by the works of Kacimi et al. in (Kacimi and Gamper, 2011; Kacimi and Gamper, 2012). We are given a set of news articles $A = \{a_1, a_2, ..., a_n\}$ where $n \geq 2$. Our goal is to select a subset $L_k \subseteq A$ of news articles that is diverse. We assume three main components that define the diversity of a set of news articles : *relevance*, *semantic diversity*, and *sentiment diversity*. Naturally, before discussing whether a set is diverse or not, it should first contain relevant news articles. Note that the *relevance* of each news article is given by the cosine similarity score as described earlier. To diversify a set of news articles, we need to give more preference to dissimilar news articles. We assume that two news articles are dissimilar if (1) they contain different tuples of entities and/or key-concepts, and/or (2) they exhibit different sentiments about those tuples. To satisfy these two requirements, we define two distance functions. The first one is a *semantic distance* function $d : A \times A \rightarrow R^+$ between news articles, where smaller the distance, the more similar the two news articles are. The second one is a *sentiment distance* function $s : A \times A \rightarrow R^+$ between news articles, where smaller the distance, the closest in sentiments the two news articles are. We formalize a set selection function $f : 2^A \times r \times d \times o \rightarrow R^+$, where we assign scores to all possible subsets of C, given a relevance function $r(.)$, a semantic distance function $d(.,.)$, a sentiment distance function $s(.,.)$, and a given integer $k \in Z^+ (k \geq 2)$. The goal is to select a set $L_k \subseteq D$ of news articles such as the value of f is maximized. In other words, the objective is to find:

$$L_k^* = Max_{L_k \subseteq D, |L_k|=k} f(L_k, r(.), d(.,.), s(.,.))$$

where all arguments other than L_k are fixed inputs to the function. The goal of this model is to maximize the sum of the relevance, the semantic dissimilarity, and the sentiment dissimilarity of the selected set. The function we aim at maximizing can be formalized as follows:

$$f(L) = \alpha(k-1) \sum_{a \in L} r(a) + 2\beta \sum_{a,b \in L} d(a,b) + 2\gamma \sum_{a,b \in L} s(a,b)$$

where $|L| = k$, and $\alpha, \beta, \gamma > 0$ are parameters specifying the trade-off between relevance, semantic diversity, and sentiment diversity[7]. The model allows to put more emphasis on relevance, on semantic diversity, on sentiment diversity, or on any mixture of these measures. Note that we need to scale up the three terms of the function. The reason is that there are $\frac{k(k-1)}{2}$ numbers in the semantic similarity sum, and $\frac{k(k-1)}{2}$ in the sentiment sum as opposed to k numbers in the relevance sum. The relevance scores are computed using cosine similarity and the semantic distance is computed using Jaccard similarity function. As for sentiment distance, we define it as follows:

$$s(a,b) = \begin{cases} 0, & \text{if the tuples have the same sentiment;} \\ 1, & \text{otherwise.} \end{cases}$$

where the sentiment orientation includes *positive*, *negative*, and *neutral* sentiments.

The problem of diversifying search results is NP-hard (Gollapudi and Sharma, 2009; Agrawal et al., 2009). However, there exist a well-known approximation algorithm to solve it (Gollapudi and Sharma, 2009), which works well in practice (Kacimi and Gamper, 2011; Kacimi and Gamper, 2012). Gollapaudi et al. (Gollapudi and Sharma, 2009) show that their Max-sum diversification objective can be approached to a facility dispersion problem, known as the MaxSumDispersion problem (Hassin et al., 1997; Korte and Hausmann, 1978). In our work, we follow the same principle and model our diversification problem as a MaxSumDispersion problem having the following objective function:

$$f'(L) = \sum_{a,b \in L} d'(a,b)$$

where d'(.,.) is a distance metric. We show in the following that f' is equivalent to our f function. To this end, we define the distance function $d'(a,b)$ as follows:

$$d'(a,b) = \begin{cases} 0, & \textit{if a=b} \\ r(a) + r(b)) + 2\beta d(a,b) & \\ +2\gamma s(a,b), & \textit{otherwise} \end{cases}$$

[7] In our implementation we have set $\alpha = \beta = \gamma = 1$

Algorithm 1: Algorithm for MaxSumDispersion.

Input: News articles C, k
Output: Set $L(|L| = k)$ that maximizes $f(L)$
Initialize the set $L = \emptyset$
for $i \leftarrow 1$ **to** $\frac{k}{2}$ **do**
  $Find(a,b) = Max_{x,y \in D} d(x,y)$
  Set $L = L \cup \{a,b\}$
  Delete all edges from E that are incident to a or b
end for
If k is odd, add an arbitrary news article to L

Considering the binary sentiment function, we claim that if d(.,.) is a metric then d'(.,.) is also a metric (proof skipped). We replace d'(.,.) by its definition in f'(L), disregarding pairwise distances between identical pairs, thus we obtain:

$$f'(L) = \alpha(k-1) \sum_{a \in L} r(a) + 2\beta \sum_{a,b \in L} d(a,b)$$

$$+2\gamma \sum_{a,b \in L} s(a,b)$$

we can easily see that each $r(a)$ is counted exactly $(k-1)$ times. Hence, the function f' is equivalent to our function f. Given this mapping, we can use a 2-approximation algorithm proposed in (Hassin et al., 1997; Korte and Hausmann, 1978) and illustrated by algorithm 1 to maximize our MaxSum objective f.

5 EXPERIMENTS

5.1 Real-world Collection

We have crawled a dataset based on the activities of a subset of 150 most active users from The Independent news site, where users follow also other news websites including The Telegraph, CNN and ALjazeera, so they have access to different types of articles with different viewpoints on the same topic. For each of those users, we have crawled their comments in the four news sites mentioned earlier from May 2010 to December 2013. Statistics about the number of comments and articles from each news web site are shown in Table 1. The distribution of users' comments and commented news articles for each user are shown in Figure 1.

5.2 Evaluation

For each user, we performed recommendation after a time point t. We used data before time point t for creating the user profile and data starting from time

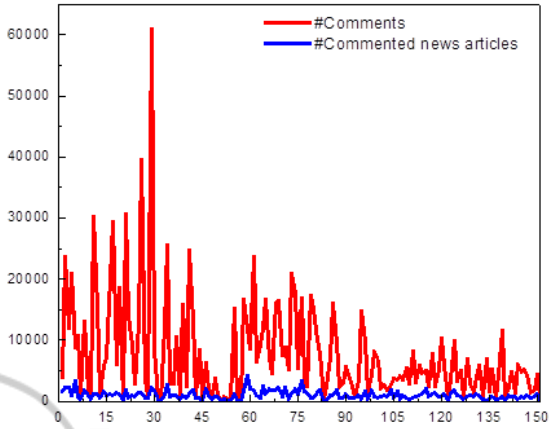

Figure 1: Users' Comments and commented News articles Distribution per user.

Table 1: Datasets Statistics.

#Comments	482, 073
#Independent articles	26, 096
#Telegraph articles	23, 154
#CNN articles	535
#Aljazeera articles	303

point t for assessment. The time point t is chosen in such a way that there is at least 200 comments posted by the user. We have used an automatic evaluation to avoid the subjectivity of manual assessments, where we consider the action of commenting on an article to be an indicator that the article fits the interests of the user. So, among the recommended articles, the ones commented by the user are considered relevant. Note that a person might well be interested in an article even though she/he does not comment on it but we did not consider that in our evaluation. As a baseline, we used the strategy proposed in (Abbar et al., 2013) where user profiles are represented by a set of entities and their related sentiments. Similarly, to the work done on (Abbar et al., 2013) we used the tool OpenCalais[8] to extract entities from news articles content and users' comments.

To compare the results of the different strategies, we use Precision and NDCG at k ($P@k$ and $NDCG@k$). The $P@k$ is the fraction of recommended articles that are relevant to the user considering only the top-k results. It is given by:

$$P@k = \frac{|Relevant\_Articles \cap topk\_Articles\_Results|}{k}$$

Additionally, we compute $NDCG$ to measure the usefulness (gain) of recommended articles based on their (geometrically weighted) positions in the result list. It

[8] www.opencalais.com

Table 2: Precision and NDCG values for all users.

	P@5	P@10	NDCG @5	NDCG @10
Entity-centric Profile (Abbar et al., 2013)	0.512	0.551	0.806	0.786
Global Profile	**0.586**	**0.593**	**0.855**	**0.797**

is computed as follows:

$$NDCG(E,k) = \frac{1}{|E|}\sum_{j=1}^{|E|} Z_{kj} \sum_{i=1}^{k} \frac{2^{rel(j,i)}-1}{log_2(1+i)}$$

where Z_{kj} is a normalization factor calculated to make *NDCG* at k equal to 1 in case of perfect ranking, and $rel(j,i)$ is the relevance score of a news article at rank i.

In our setting, relevance scores $rel(j,i)$ have two different values: 1(relevant) if the news article was commented by the user u, and 0(not relevant) if the news article was not commented by the user u. The precision and *NDCG* results for the three strategies are shown in Table 1.

We can see in Table 1 that our approach of using global profile outperforms the baseline approach with a gain between 4 and 7 of % in term of precision and 5% in term of ranking at *NDCG*@5. The reason is that most of news articles do not address entities without relating them to some key-concepts. Moreover, when viewpoints are expressed about entities, they usually refer to certain key-concepts of those entities. Thus, using only entities to build profiles gives less room for diversification which penalizes the performance. Consequently the combination of both entities and key-concepts give the best results.

6 CONCLUSION AND OUTLOOK

In this paper, we have proposed a two-stage personalized news recommendation approach that takes into account users interests based on their comments in news sites. We recommend a set of diverse news articles using dissimilarity measure based on (1) semantic diversification and/or (3) sentiment diversification. As future works, we plan to first test our model in a bigger set of users and explore more diversification techniques based on users' comments.

REFERENCES

Abbar, S., Amer-Yahia, S., Indyk, P., and Mahabadi, S. (2013). Real-time recommendation of diverse related articles. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, pages 1–12, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

Abel, F., Gao, Q., Houben, G.-J., and Tao, K. (2011). Analyzing user modeling on twitter for personalized news recommendations. In *Proceedings of the 19th International Conference on User Modeling, Adaption, and Personalization*, UMAP'11, pages 1–12, Berlin, Heidelberg. Springer-Verlag.

Agrawal, R., Gollapudi, S., Halverson, A., and Ieong, S. (2009). Diversifying search results. In *WSDM*, pages 5–14.

Carterette, B. and Chandar, P. (2009). Probabilistic models of ranking novel documents for faceted topic retrieval. In *CIKM*, pages 1287–1296.

Chen, J., Nairn, R., Nelson, L., Bernstein, M., and Chi, E. (2010). Short and tweet: Experiments on recommending content from information streams. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 1185–1194, New York, NY, USA. ACM.

Clarke, C. L. A., Kolla, M., Cormack, G. V., Vechtomova, O., Ashkan, A., Büttcher, S., and MacKinnon, I. (2008). Novelty and diversity in information retrieval evaluation. In *SIGIR*, pages 659–666.

Gollapudi, S. and Sharma, A. (2009). An axiomatic approach for result diversification. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, pages 381–390, New York, NY, USA. ACM.

Hassin, R., Rubinstein, S., and Tamir, A. (1997). Approximation algorithms for maximum dispersion. *Operations Research Letters*, 21:133–137.

Hong, L. and Davison, B. D. (2010). Empirical study of topic modeling in twitter. In *Proceedings of the First Workshop on Social Media Analytics*, SOMA '10, pages 80–88, New York, NY, USA. ACM.

Kacimi, M. and Gamper, J. (2011). Diversifying search results of controversial queries. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, CIKM '11, pages 93–98, New York, NY, USA. ACM.

Kacimi, M. and Gamper, J. (2012). Mouna: Mining opinions to unveil neglected arguments. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, CIKM '12, pages 2722–2724, New York, NY, USA. ACM.

Korte, B. and Hausmann, D. (1978). An analysis of the greedy heuristic for independence systems. *Annals of Discrete Mathematics*, 2:65–74.

Li, Q., Wang, J., Chen, Y. P., and Lin, Z. (2010). User comments for news recommendation in forum-based social media. *Inf. Sci.*, 180(24):4929–4939.

Meguebli, B.-L. Y., Kacimi, M., Doan, B.-l., and Popineau, F. (2014a). Building rich user profiles for personal-

ized news recommendation. In *Proceedings of 2nd International Workshop on News Recommendation and Analytics*.

Meguebli, Y., Kacimi, M., Doan, B.-L., and Popineau, F. (2014b). Unsupervised approach for identifying users political orientations. In *Advances in Information Retrieval*, pages 507–512. Springer.

Michelson, M. and Macskassy, S. A. (2010). Discovering users' topics of interest on twitter: A first look. In *Proceedings of the Fourth Workshop on Analytics for Noisy Unstructured Text Data*, AND '10, pages 73–80, New York, NY, USA. ACM.

Phelan, O., McCarthy, K., and Smyth, B. (2009). Using twitter to recommend real-time topical news. In *Proceedings of the Third ACM Conference on Recommender Systems*, RecSys '09, pages 385–388, New York, NY, USA. ACM.

Radlinski, F. and Dumais, S. T. (2006). Improving personalized web search using result diversification. In *SIGIR*, pages 691–692.

Santos, R. L. T., Macdonald, C., and Ounis, I. (2010). Selectively diversifying web search results. In *CIKM*, pages 1179–1188.

Shmueli, E., Kagian, A., Koren, Y., and Lempel, R. (2012). Care to comment?: Recommendations for commenting on news stories. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, pages 429–438, New York, NY, USA. ACM.

Singhal, A. (2001). Modern information retrieval: a brief overview. *BULLETIN OF THE IEEE COMPUTER SOCIETY TECHNICAL COMMITTEE ON DATA ENGINEERING*, 24:2001.

Stoyanovich, J., Amer-yahia, S., Marlow, C., and Yu, C. (2008). Leveraging tagging to model user interests in del.icio.us. In *In AAAI SIP*.

Wang, J. and Zhu, J. (2009). Portfolio theory of information retrieval. In *SIGIR*, pages 115–122.

Weng, J., Lim, E.-P., Jiang, J., and He, Q. (2010). Twitterrank: Finding topic-sensitive influential twitterers. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, WSDM '10, pages 261–270, New York, NY, USA. ACM.

Zhai, C. and Lafferty, J. D. (2006). A risk minimization framework for information retrieval. *Inf. Process. Manage.*, 42(1):31–55.