World Wide Web
DOI 10.1007/s11280-017-0436-2

Towards better news article recommendation

With the help of user comments

Youssef Meguebli[1] · Mouna Kacimi[2] ·
Bich-Liên Doan[3] · Fabrice Popineau[3]

Received: 20 June 2016 / Revised: 15 November 2016 / Accepted: 11 January 2017

Abstract News media platforms publish articles about daily events letting their users comment on them, and forming interesting discussions in almost real-time. To keep users always active and interested, media platforms need an effective recommender system to bring up new articles that match user interests. In this article, we show that we can improve the quality of recommendation by exploiting valuable information provided by user comments. This information reveals aspects not directly tackled by the news article on which they have been posted. We call such aspects *latent aspects*. We demonstrate how these latent aspects can make a crucial difference in the accuracy of future recommendation. The challenge in detecting them is due to the noisy nature of user comments. To support our claim, we propose a novel news recommendation system that (1) enriches the description of news articles by latent aspects extracted from user comments, (2) deals with noisy comments by proposing a model for user comments ranking, and (3) proposes a diversification model to remove redundancies and provide a wide coverage of aspects. We have tested our approach using large collections of real user activities in four news Web sites, namely The INDEPENDENT, The Telegraph, CNN and Al-Jazeera. The results show that our approach outperforms baseline approaches achieving a significantly higher accuracy.

✉ Mouna Kacimi
mouna.kacimi@unibz.it

Youssef Meguebli
meguebli.youssef@orange.fr

Bich-Liên Doan
bich-lien.doan@centralesupelec.fr

Fabrice Popineau
fabrice.popineau@centralesupelec.fr

1 Orange Labs – France Telecom, 4 rue du Clos Courtel, Rennes, France

2 Free University of Bozen-Bolzan, Piazza Domenicani 3, 39100 Bozen-Bolzano, Italy

3 LRI – CentraleSupélec, 91405 Saclay, France

Published online: 28 February 2017

Keywords News recommendation · User comments ranking · Diversification

1 Introduction

Media platforms, like CNN[1] and The INDEPENDENT,[2] continuously deliver the latest breaking news on various topics. Moreover, they provide users the possibility to write comments about any published article and engage in discussions with other users. Figure 1 shows an example of a news article about Donald Trump's victory speech, published on The INDEPENDENT at March 02, 2016. The article talks about *"Donald Trump winning 7 states out of 12"* and *"his speech being polite, cordial and humble"*. As a reaction, users posted comments to discuss what was reported in the news. These comments do not focus only on the speech of Donald Trump but discuss also other aspects such as *"Foreign Policy"*, *"Hilary Clinton's chances of winning"*, *"anti-establishment"*, *"US nomination process"*, and *"Donald Trump's wealth"*.

To provide users with such insightful discussions requires first finding news articles that match user interests. This task can be very tedious when users are overwhelmed by a huge number of news articles. Therefore, a key challenge for media platforms is to keep the attention of users by offering an effective news recommendation service. Typically, both users and news articles are represented by sets of features. In the context of media platforms, user features are extracted from the comments he posts. By contrast, news article features are extracted from its content. Based on this representation, a news article is recommended to a user if their features are similar. The problem of applying this standard representation is the gap that often exists between the content of the news article and the comments of users. Looking at the previous example, the news article is about Donald Trump's victory speech while some of its comments discuss *'Foreign Policy"* or *"Hilary Clinton's chances of winning"*. Users posting those comments would have very different features from those of the news article and thus they would not be considered for recommendation, which is obviously misleading.

To overcome the above problem, it is important to measure the interestingness of news articles on more information than the one available in their mere content. What makes an article interesting are not only the events it reports, but the whole context into which these events happen including related past events or future consequences. For this reason, users often express their interests in an article by discussing aspects which are not explicitly mentioned in the news article. We call these aspects *latent aspects*. Although these latent aspects are not the main subject of the news article, they represent strong indicators of what makes the news article interesting for users. Hence, a natural way to enrich news article features is by integrating the latent aspects appearing in user comments.

The core problem of our work is how to extract latent aspects from user comments for enriching news article contents and improving recommendation. Several techniques employing user comments have been proposed in the literature [1, 7, 30, 37, 37], including collaborative and content filtering techniques. Most of these approaches are related to product reviews while very few focus on news article recommendation. Particularly, they address the problem of real-time recommendation tackling mainly diversification problems [1]. Exploiting existing approaches for our purpose brings new challenges: (1) aspects about

[1] www.cnn.com

[2] www.independent.co.uk

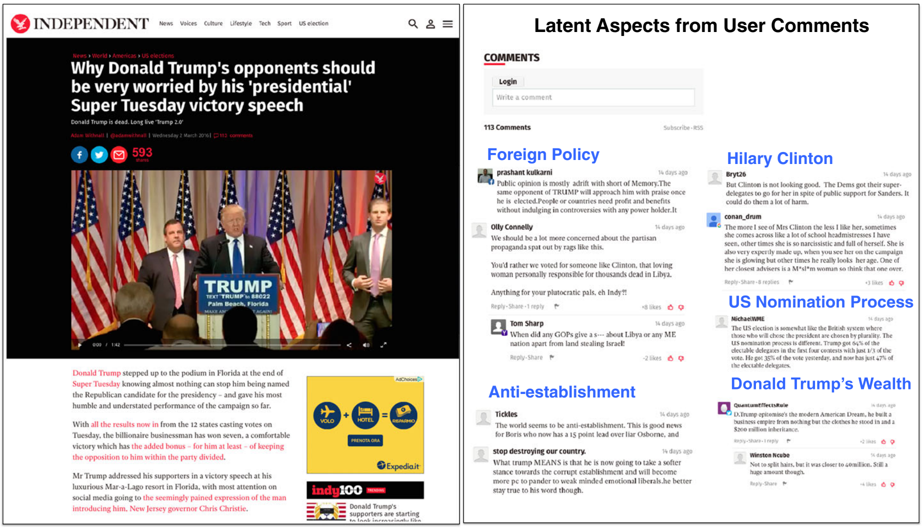

Figure 1 A news article and its comments

news events are much more challenging to extract compared to product reviews applications, where aspects are well defined and pre-classified (2) user comments are informal and might include redundancies and noisy information such as abbreviations and misspellings, which make aspects extraction difficult (3) discussions of users evolve quickly which leads to topic drift problems. This makes the selection of user comments to be used for enriching news article contents a challenging task.

In this article, we propose an approach for aspect extraction from news articles and user comments on media platforms, tackling the problems above. The goal is to enrich news articles and improve recommendation. The main contributions of this article are summarized as follows:

1. We propose an aspect extraction approach for both news article contents and user comments. The latent aspects extracted from user comments are then used to enrich the content of news articles.
2. We propose a scoring model for user comments using PageRank algorithm, to address noise and topic drift problems
3. We propose a diversification model for user comments to handle the problem of redundancies and cover more aspects related to the news article content
4. We evaluate our aspect extraction in a recommendation scenario using large collections of real user activities in four news Web sites, namely The INDEPENDENT, The Telegraph, CNN and Al-Jazeera. The results show that our approach highly increases precision compared to baseline approaches

The rest of this paper is organized as follows. Section 2 presents related work. Section 3 introduces the problem we are tackling in this work. Section 4 describes two strategies to select comments for enriching news articles profiles. Section 5 presents our approach for aspect extraction from news content and user comments. Section 6 presents and discusses experiment results, and finally Section 7 concludes the article.

2 Related work

User comments have been exploited in the literature for different purposes including blog summarization [11], community detection [28], spam detection [12, 27, 36], opinion volume and rating prediction [17, 22, 24, 32], opinion ranking [18, 33], opinion summarization [3, 19, 21, 38], and the identification of user political orientations [23]. Directly related approaches to our work exploit user comments for search and recommendation [1, 4, 7, 15, 29, 30, 37, 37]. Li et al. [15] exploit user comments for detecting topic evolution and hence improving recommendation. San Pedro et al. [29] use user comments for re-ranking image search results. Ganesan and Zhai [7] extract entities from user comments with the goal of recommending to users only comments that contain entities of their interests. Shmueli et al. [30] analyze co-commenting patterns of users for recommending news articles to users who will likely comment them. Similarly, Bansal et al. [4] recommend comment-worthy articles by establishing relations between user comments and news article topics. Abbar et al. [1] address the problem of online recommendation where they extract from user comments a set of features to perform news articles diversification. The closest work to ours is by, Yee et al. [37] employing user comments to enrich the content of documents. They prove that the incorporation of Youtube user comments in the search index result in an up to 15 % improvement of search accuracy compared to user-supplied tags or video titles. The difference with our work is that (1) we deal with the domain of news recommendation, and (2) we exploit only highly ranked user comments due to their noisy nature. Moreover, we perform diversification on those comments to have a broader coverage.

Enriching news articles with user comments requires first the extraction of useful knowledge from them. Following this direction of research, several approaches [26, 35, 39] have focused on aspect extraction from annotated data. For instance, H. Wang et al. [35] identify the main aspects of reviews by starting from few seed keywords which are fed into a bootstrapping-based algorithm. Zhuang et al. [39] propose a multi-knowledge based-approach which integrates WordNet, statistical analysis, and movie knowledge to extract a list of features that users have commented on and create a review summary. Popescu and Etzioni [26] describe an unsupervised information-extraction system for product features reviewed by users. Most of these approaches are either domain-specific or highly dependent on training data. Our work share the same goal of aspect extraction from user comments, however we do not focus on product reviews but address general topics on daily events. In addition, we employ an unsupervised approach for a more flexible solution.

A second key point when dealing with user comments is how to select only relevant ones for a given task. To address this issue, several approaches proposed user comment ranking strategies where most of them are for product review domain. Kim et al. [13] train an SVM regression system to learn a helpfulness function used to rank user comments. Using extensive experiments, Danescu-Niculescu-Mizil et al. [6] show that exploiting relationships between user comments can significantly improve ranking quality. Tsur and Rappoport [34] use a lexicon of dominant terms to create a feature vector representation for user comments which are then ranked based on their distance from a core of a virtual optimal comment. Liu et al. [20] propose a nonlinear regression model for comment helpfulness prediction based on user expertise, comment writing style, and its timeliness. The closest work to ours is by Litvak and Matz [18], dealing with user comments ranking in news media platforms. The authors propose to rank user comments based on their relevance to a given paragraph using PageRank. They build a graph of user comments where an edge is created if the cosine similarity between two comments is higher than a given threshold. The difference with our work is that we take answers as relations between user comments, thus giving more weight to comments that trigger more discussions.

A directly related area to our work is recommendation systems. Several approaches have been proposed [1, 2, 5, 16, 25, 30] where two main strategies have been adopted and mostly combined. First, content filtering strategies which create a profile for each user or seed article and then recommend the best matching articles based on user profile, seed article, or both. Second, collaborative filtering strategies which 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.

3 Problem definition

We are in the scenario where we aim at recommending to users news articles of their interests. To this end, we need to have effective representations of users and news articles. The representations we adopt are based on the notion of *Aspects*. We define an aspect as any subject of discussion present in the content of news articles or user comments. An aspect can be an entity such as a person, an organization, a location, or any well-defined concept such as languages, nationalities, or wars. Additionally, it can be a matter related to an entity such as possessions, ideas, or actions. Consider the following user comment:

> *"Trump will make America believe again; Hillary cannot do that. He will win, he will smash the left, he will recover American self-confidence. Europe and Britain will hate him and that will be a bonus."*

Example of aspects discussed by the above comment are: *Trump*, *Hilary*, *American self confidence*, *Europe*, and *Britain*.

We distinguish two types of aspects: *News Aspects* and *Latent Aspects*. The former are the aspects present in the news article and the latter are the aspects discussed in user comments but not explicitly mentioned or clearly described in the news article. Latent aspects are used to enrich the set of news aspects since they (1) reflect what triggers the attention of users when reading the article (2) show the relatedness of the article to other events, which plays an important role to target the right users for recommendation. In the above example, *Trump*, *Hilary* are news aspects while *American self confidence*, *Europe*, and *Britain* are latent aspects because they are raised by user discussions. Such aspects are used to build news article and user profiles as described in the following:

News Article Profile each news article is described by a vector of weighted aspects $\langle(a_1, w_1), (a_2, w_2), ..., (a_r, w_r)\rangle$. These aspects are extracted from the content of the news article. Note that the content of the news article can be enriched using user comments as described later. Thus, the aspects composing the news article profile can be a mixture of news aspects and latent aspects.

User Profile each user of the news platform is described by a vector of weighted aspects $\langle(a_1, w_1), (a_2, w_2), ..., (a_u, w_u)\rangle$. These aspects are extracted from all the comments posted by the user in the news platform. They can refer to both news aspects and latent aspects depending on whether the user is discussing only what is reported by the news article or giving new insights.

To perform recommendation, we use a content filtering approach where we compare the profiles of news articles to a given user profile. The most similar news articles are then recommended to the corresponding user. We use the above profiles in a vector space model to compute the similarity between users and news articles based on cosine similarity, an effective technique for text-based applications.

It is important to note that our contribution focuses on building news article and user profiles based on latent aspects. Consequently, the key problem is to develop an effective technique for aspect extraction from unstructured text. We can clearly see that the aspects mentioned in the previous example, including *Trump*, *Hilary*, *American self confidence*, *Europe*, and *Britain*, can be extracted in a straightforward way considering only entities and noun phrases. However, the comment discusses also the following ideas: *Trump winning and smashing the left*, *Hilary loosing*, and *Europe and Britain hating Trump*. These later aspects are certainly interesting but also much more challenging to extract due to the complexity of natural language processing.

Another key problem is related to the enrichment of news article profiles using latent aspects. Latent aspects are extracted from user comments. However, the number of user comments can highly increase especially when the news article reports controversial topics that open the door to various opinions and arguments. Some of these comments are related to the news and generate many discussions, however some others are either off topic or they do not capture the attention of other users. Thus, extracting latent aspects from all user comments is subject to noise and inconsistencies which would have a negative impact on the quality of recommendation. To solve this problem, we aim in this work at finding the set of user comments that discuss latent aspects (1) related to news aspects avoiding topic-drift problems, and (2) shared among a large number of users to avoid overfitting.

To find out which user comments we have to use for enriching news aspects, it is important to analyze their structure. As shown in Figure 2, users post their comments following a tree structure. Generally, the full tree has a news article as a root, and comments as either intermediate nodes or leaves. An edge represents a reaction relationship. A comment node is either reacting directly to the content of the news article and thus is directly related to the root, or is answering another comment represented as its parent in the tree. Based on

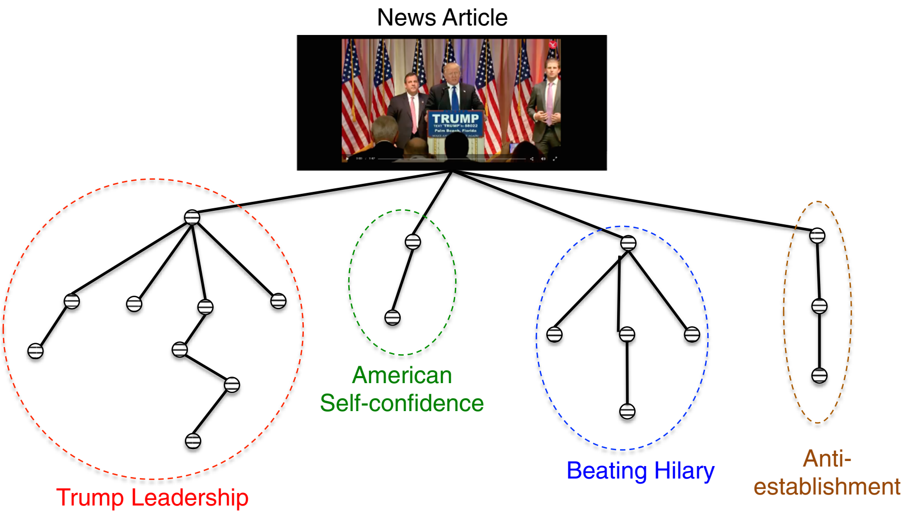

Figure 2 The structure of user comments

this structure, the different sub-trees correspond to the various discussions around the news article. In the example shown in Figure 2, comments discuss *Trump Leadership*, *American Self-confidence*, *Beating Hilary*, and *Anti-establishment*. We can see here that not all these aspects are of the same importance. *Trump Leadership* and *Beating Hilary* discussions involve more comments and interactions between users than *American Self-confidence* and *Anti-establishment*. Therefore, our goal is to exploit the link structure to rank comments according to their importance.

Ranking comments reduces the amount of noise and improves the quality of latent aspects. However, it does not solve the problem of redundancies. Many users repeat what was reported by other comments expressing the same ideas in different ways. Changing the writing style does not mean having more latent aspects. For this reason, it is very important to diversify the set of aspects extracted from user comments. Let us take the following examples:

Comment1

> *"I don't think he believes anything Trump says. He says what he thinks you want to hear, that is all."*

Comment2

> *"Trump is saying what he needs in order to get elected but once in he would more than likely take a middle path but with a bit less brown nosing."*

Comment3

> *"Trump says what he thinks and many agree. The left say what they want people to think."*

All the above comments talk about what *Trump says*. We note that these comments talk about the same aspect but in very different ways using different vocabulary. Thus, applying diversification using the whole content of comments is misleading since these comments are very dissimilar in terms of common words. To tackle this problem, we propose in this article a diversification model that finds a diverse set of comments that lead to a diverse set of latent aspects independently from their writing styles.

4 User comments selection

4.1 Ranking model

The first step towards our goal is to find the set of user comments from which latent aspects are extracted to enrich news articles. The reason for not taking all user comments is that some comments can be noisy or completely unrelated to the news article. To solve this problem, our aim is to propose a ranking model for user comments, where highly ranked comments would guarantee an effective enrichment of news aspects. As we can observe from the way users post their comments, there exist several threads of discussions. Each thread focuses on certain aspects. Moreover, some threads contain much more comments than others. Intuitively, the more discussions around a given aspect, the more interesting is the aspect and the more related it is to the news article. Based on this intuition, we build our ranking model by exploiting user interactions.

As described in Section 3, news articles and their comments have a tree structure. More precisely, they have an in-tree structure where all the links point towards the root as shown in Figure 3. On one hand, a comment can have only one outgoing link because it can answer only one comment. On the other hand, it can have more than one incoming link when it gets more than one answer. A comment can be posted directly on the news article and thus does not have any outgoing link to another comment. It can also be a leaf, meaning that it does not get any answer and thus does not have any incoming links. In our work, we exploit this link structure to compute comment scores. The idea is to compute authority scores for user comments using the PageRank Algorithm. PageRank considers links as votes. Comments with more incoming links are more important, and links from important comments are also more important.

To compute authority scores, we model user behavior in the comments graph, where a random user visits a comment with a certain probability based on the comment authority score. The probability that the random user clicks on one link is solely given by the number of links on that comment. So, the probability of the random user to reach one comment is the sum of probabilities for the random user to follow links to this comment. It is assumed that the user does not click on an infinite number of links, but gets bored and jumps to another comment at random. Besides its interpretation, the random jump is used to avoid dead-ends which correspond to comments with no outgoing links. Note that the random jump is also used to avoid spider traps, but this does not apply to the comments graph since it cannot contain loops. Formally, the authority score is given by:

$$h(C) = (1 - m) + m\left(\frac{h(T_1)}{L(T_1)} + \ldots + \frac{h(T_n)}{L(T_n)}\right)$$

where $h(C)$ is the authority score of comment C, $h(T_i)$ is the authority score of comments T_i which link to comment C, $L(T_i)$ is the number of outgoing links of comment T_i, and m

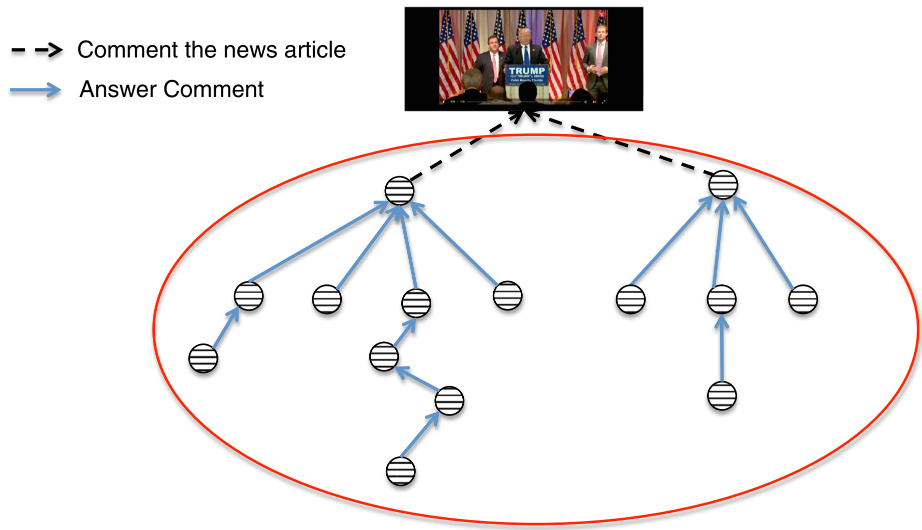

Figure 3 User comments graph

is a damping factor which can be set between 0 and 1. As we can see, the authority score of comment C is recursively defined by the authority scores of comments which are linked to it. The authority score of a comment T is always weighted by the number of its outgoing links. The weighted authority scores of comments T_i are then added up. Finally, the sum of the weighted authority scores of all comments T_i is multiplied with a damping factor m which reflects the probability for the random surfer not stopping to click on links.

4.2 Diversification model

After defining a ranking model, we can select the top-k user comments to be used for latent aspect extraction. However, redundancy occurs often specially in debates where users try to make clear their view points. They might use different arguments or ways to express their ideas but they still talk about the same aspects. For this reason, taking the top-k user comments relying only on authority scores does not guarantee a large coverage of latent aspects. To solve tis problem, we extend our ranking model using diversification.

We formulate the diversification problem as follows. We are given a set of user comments $C = \{c_1, c_2, ..., c_n\}$ where $n \geq 2$. Our goal is to select a subset $S_k \subseteq C$ of user comments that is diverse. We assume two main components that define the diversity of a set of user comments : *authority* and *semantic diversity*. Naturally, before discussing whether a set is diverse or not, it should first contain user comments with high authority scores. Among user comments having high authority scores, we need to give more preference to those which are dissimilar. We assume that two comments are dissimilar if they discuss different latent aspects. To satisfy this requirement, we define a *semantic distance* function $d : C \times C \rightarrow R^+$ between user comments, where the smaller the distance, the more similar the two user comments are. This distance measures the *semantic diversity* of the set. Using the diversification framework proposed in Gollapudi and Sharma [8], we formalize a set selection function $f : 2^C \times h \times d \rightarrow R^+$, where we assign scores to all possible subsets of C, given an authority function $h(.)$, and a semantic distance function $d(.,.)$, and a given integer $k \in Z^+ (k \geq 2)$. The goal is to select a set $L_k \subseteq D$ of user comments such as the value of f is maximized. In other words, the objective is to find:

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

where all arguments other than L_k are fixed inputs to the function.

The goal of this model is to maximize the sum of authority and semantic dissimilarity for a selected comment set. The function we aim at maximizing can be formalized as follows:

$$f(L) = (k-1) \sum_{a \in L} h(a) + 2\lambda \sum_{a,b \in L} d(a,b)$$

where $|L| = k$, and $\lambda > 0$ is a parameter specifying the trade-off between authority and semantic diversity.[3] The model allows to put more emphasis on authority, semantic diversity, or a mixture of them. Note that we need to scale up the first sum to balance out the fact that there are $\frac{k(k-1)}{2}$ numbers in the similarity sum, as opposed to k numbers in the authority sum.

The problem of diversifying search results is NP-hard as shown by Gollapudi and Sharma [8]. However, there exists a well-known approximation algorithm to solve it, which works well in practice Gollapudi and Sharma [9] show that their Max-sum diversification objective

[3]We have experimentally set $\lambda = 0.5$

can be approached to as a facility dispersion problem, known as the MaxSumDispersion problem[10, 14]. In our work, we follow the same principle and model our diversification problem as a MaxSumDispersion problem that has 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:

$$d'(a,b) = h(a) + h(b) + 2\lambda d(a,b)$$

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:

$$f'(L) = \sum_{a,b \in L} (h(a) + h(b)) + 2\lambda \sum_{a,b \in L} d(a,b)$$

We can easily see that each $h(a)$ is counted exactly $(k-1)$ times (considering the complete graph on L) . Hence, the function f' is equivalent to our function f. Given this mapping, we can use a 2-approximation algorithm as proposed in Hassin et al. [10]. This is illustrated by algorithm 1 to maximize our MaxSum objective *f*.

Algorithm 1 Algorithm for MaxSumDispersion

Input: Comments 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 comment to L

5 Aspect extraction

Aspects are extracted from news content and user comments to build user and news article profiles, as described in Section 3. In our work, we employ a phrase-aware model to identify aspects. We practically construct phrases from the bag of words composing the input text. The main intuition behind is that news articles and user comments are provided by different journalists and users who have very different writing styles. So, using a model that depends on the structure of the input text, such as n-grams, would not allow us to identify aspects expressed in various ways. For example, *Trump international policy* and *international policy of Donald Trump* indicate exactly the same aspect, however they have different structures. The phrase-aware model helps to solve this problem by (1) capturing aspects independently from their exact form in the input text, and (2) extracting aspects of various sizes and types including proper names, noun phrases, and any other form.

The aspect extraction process is composed of two steps. First, we extract the set of words composing the input text after eliminating stop words. The extracted words represent the source from which the phrases are constructed. For an efficient processing, we can limit the set only to frequent words. Second, we construct phrases of various sizes. These phrases are then weighted based on the correlation of their contained words to give more importance to meaningful phrases. Then, highly ranked phrases are selected as aspects.

5.1 Phrase mining

For a given input text, we extract all its contained words. The input text can be either the content of a news article (enriched by a subset of top-k ranked or diversified user comments) or all the comments of a user, depending on whether we build a news article profile or a user profile. After eliminating stop words, we rank all remaining words based on their frequency scores. This ranking aims at reducing the processing cost of the next steps. We weight each word based on the $tf - idf$ model as described in the following:

$$tf - idf(w_i) = tf(w_i) \times idf(w_i)$$

where, $tf(w_i)$ and $idf(w_i)$ are the term frequency and the inverted document frequency of word w_i, respectively. If word w_i is extracted from a news article content, then $tf(w_i)$ represents the frequency of word w_i in the article, and $idf(w_i)$ represents the inverted document frequency considering all articles in the news website. By contrast, if the word is extracted from the comments of a given user, then $tf(w_i)$ represents the frequency of word w_i in the set of comments of that user, and $idf(w_i)$ represents the inverted document frequency considering all user comments in the news website.

After extracting the set of words, we construct phrases of various sizes. In theory, we can construct phrases of any size. However, in our application, all phrases of size greater than three were not meaningful. Of course, this can be different for other applications, so it is enough to increase or reduce the size of generated phrases.

5.2 Weighting model

From the way we generated phrases, it is clear that not all of them would have a meaning. To select the most meaningful phrases, it is important for their contained words to be strongly associated in at least one sentence of the input text. To capture this association, we use *pointwise mutual information* [31] (*PMI*) of words. Formally, suppose $s_k = w_1 ... w_n$ is a generated phrase. We define the *PMI* score as follows:

$$PMI(s_k) = \frac{1}{n} \sum_{i=1}^{n} pmi_{local}(w_i) \tag{1}$$

where n is the length of phrase s_k and $pmi_{local}(w_i)$ is a local *pointwise mutual information* function defined as:

$$pmi_{local}(w_i) = \frac{1}{2C} \sum_{j=i=C}^{i+C} pmi'(w_i, w_j), i \neq j \tag{2}$$

where C is a context window size. Note the context window size indicates the number of right and left neighboring words, to a given word w_i, that needs to be taken into account. For example a context window of size 2 means that for each word w_i we consider the two words on its right and the two on its left within the text. The $pmi_{local}(w_i)$ measures the average association strength of word w_i with all its C neighboring words (on the left and right). For example, in the phrase *Trump International policy*, assuming $C = 1$, we would first obtain the average *PMI* score of *Trump* associated with *international* and for *international* we would then obtain the average *PMI* of *international* associated with *Trump* and at last *international* with *policy*.

We use a modified *PMI* scoring (like in [7]) referred to as *pmi*′ where the *pmi*′ between two words, w_i and w_j, is defined as:

$$pmi'(w_i, w_j) = \log_2 \frac{p(w_i, w_j) \cdot c(w_i, w_j)}{p(w_i) \cdot p(w_j)} \tag{3}$$

where $c(w_i, w_j)$ is the frequency of two words co-occurring in a sentence taken from the input text within the context window of C (in any direction), and $p(w_i, w_j)$ is the corresponding joint probability. The co-occurrence frequency, $c(w_i, w_j)$, which is not part of the original *PMI* formula, is integrated into our *PMI* scoring to reward frequently occurring words in the input text. By adding $c(w_i, w_j)$ into the *PMI* scoring, we ensure that low frequency words do not dominate and moderately associated words with high co-occurrences have fairly high scores. Table 1 shows some examples of phrases extracted from the Donald Trump's news article together with its related comments.

The extracted phrases are then ranked based on their PMI scores. The topk phrases are chosen to represent aspects. Note that the number of selected phrases depends on the input text and the application. It can be defined experimentally by choosing the value of k that provides the highest quality of recommendation and user satisfaction.

6 Experiments

6.1 Datasets

We have crawled four news websites, namely, CNN, The Telegraph, The INDEPENDENT, and Al-Jazeera in a time period going from May 2010 to December 2013. We have chosen these four news websites for the following reasons. First, they share mostly the same topics including national, and international events, sports, business, culture, entertainment, which helps covering a wide range of news from different sources. Second, they are socially, culturally and sometimes politically different. This diversity results in a large number of latent aspects providing an excellent test data for our approach. Third, they all use the same

Table 1 Example of generated aspects

Example of Generated Aspects
Trump, Clinton, Vote, Supporters, Business, Election, Establishment,
President, crazies,Stupid Trump, Good Trump, Crazy Trump, Sbourne,
Trump,Trump electable, Trump International, Trump favourite, Supreme
Trump, Donald Trump vote, Donald Trump supporters, Trump vote suddenly,
Donald Trump business, Donald Trump Establishment, Donald Trump president,
Donald Trump crazies, Donald Trump trickster, Trump international policy

infrastructure for publishing user comments. In this way, we could benefit from the activities of users who are members of all four websites, which is crucial for performing cross validation.

To assess the effectiveness of news recommendation, it is important to have complete information about news articles and user activities during a given period of time. First, for each news article, we need to have information about all its related comments. This is crucial to build its profile considering latent aspects provided by all users who posted comments on it. Second, for each user we need to know all the news articles he read and the comments he posted. Practically, it was not possible to get information about the articles users clicked on, or the time they spent reading a given article. The only way to know whether a user have read an article and was interested in it is by looking at the articles he commented on. Of course that would give us a subset of the real activities, however we are sure that this subset is accurate.

To collect such information, we have chosen users as seeds for our crawler, hence we guarantee having complete information about each user. The choice of seed users was made as follows. A seed user has to satisfy three criteria. First, he should be active. We have defined an active user as the one who has published comments on at least 500 different news articles. Second, he should have continuous and sustained activity over an extensive period of time. We have set this period to 15 months. Third, he should have published comments on the four news websites. These criteria were set to ensure the feasibility of having enough history information to build user profiles and of predicting future activities. Moreover, to show that our evaluation process is independent from a specific platform. Note that the thresholds we have chosen depend on the type and quantity of data corresponding to the period from May 2010 to December 2013.

We have started our crawling by selecting the 500 most active users from each news Web site reaching a total of 2000 users. After applying the three criteria mentioned above, we had at the end 233 distinct seed users. More precisely, 150 seed users from CNN, 180 seed users from The Telegraph, 164 seed users from The INDEPENDENT, and 151 seed users from Al-Jazeera. Naturally, the same user can be a seed in more than one website which explains the above numbers. For each seed user, we have collected all the news articles he commented on and the list of all the comments he posted. For each news article, we have collected all the comments related to it. The result of this process was a large dataset of news articles and user comments. More detailed statistics about the dataset are shown in Table 2.

Table 2 Dataset statistics

	#articles	*#comments*	*#articles*	*#comments*
	Seeds from CNN		Seeds from The Telegraph	
CNN	41, 245	12, 056, 789	665	874, 879
Telegraph	1, 908	1, 257, 645	56, 527	10, 704, 741
Independent	1, 412	987, 437	7, 999	1, 608, 665
Al-Jazeera	801	102, 254	451	62, 835
	Seeds from The INDEPENDENT		Seeds from Al-Jazeera	
CNN	528	421, 542	2, 233	1, 652, 875
Telegraph	23, 272	6, 710, 580	1, 126	894, 710
Independent	27, 012	2, 985, 412	394	54, 760
Al-Jazeera	303	48, 058	9, 313	531, 452

6.2 Evaluation

We have used two baseline approaches to show the impact of our approach. Additionally, we have tested two ways of selecting user comments to enrich news articles profiles. So we have four strategies under comparison which are:

NoEnrich The first baseline is a simple content filtering approach based solely on the news article content, meaning that no enrichment of articles with comments was performed.

AllComments The second baseline is a closely related work [37] to ours. This baseline enriches each news article with its whole set of comments.

Authority We use our approach to enrich news articles with the top-k authoritative comments related to it. The top-k comments are selected using the strategy described in Section 4.1. In our experiments we used several values of k and the best results corresponded to $k = 10$.

Diversity We use our approach to enrich news articles with the most diverse top-k comments related to it. Diversification is performed as described in Section 4.2. In our experiments we used several values of k and the best results corresponded to $k = 10$.

These baselines have been chosen to show the importance of comment selection to enrich news articles profiles. Especially the **AllComments** baseline is a very strong and valid baseline for our approach. Indeed, we show that selecting all comments to enrich news articles profiles is not optimal and that our comment selection strategy improves significantly the effectiveness of recommendation.

For each user, we have performed a recommendation at different time points $t_1, t_2, ..., t_n$. The reason for using time dependent evaluation is twofold. First, to take into account profile updates since users continuously post comments bringing new information about their interests. Second to use data before time point t_i for recommendation and data starting from time point t_i for assessment, as described below. The time points $t_1, t_2, ..., t_n$ are chosen in such a way that between t_{i-1} and t_i, there is at least m news articles commented by the user. For each user u_i, we have chosen $m = \frac{N_i}{10}$ where N_i is the total number of commented news articles by the user u_i. This setting resulted in 2330 rounds of recommendation.

In order to avoid the subjectivity of manual assessment, we do an automatic assessment based on the data we have collected. We have information about all the news articles and the users who commented on them. So, if a user comments on an article, we consider the article relevant to him. For each user we have the full history of his comments and the news articles that he commented on. So for performing experiments, we break this data into two parts:

– Part1: before time point t, the data is used to build user profile
– Part2: after time point t, the data is used for recommendation

The articles recommended to the user are from Part2. For each of those articles we know whether the user have commented on them or not since we have the full history as described earlier. Therefore, when we check the top-k articles recommended to a given user, the number of relevant articles is simply the number of articles he commented on them.

Based on these assumptions, we check the list of recommended news articles among the 2330 rounds of recommendation. Note that some information is likely to be missing: a person might well be interested in an article even though he/she does not comment on it. So, the real results are most probably more effective than our findings.

We assess the effectiveness of our approach using Precision at n (P@n), Recall at n (R@n) and the F1 score. The P@n is the fraction of articles among the top-n results relevant to a given user over the top-n results. The R@n is the fraction of articles among the top-n results relevant to a given user over the total number of articles relevant to this user. The F1 score is defined as twice the harmonic mean of these two values and combines both P@n and R@n in a single score value comprised between 0 and 1, 1 being the perfect score.

$$P@n = \frac{|\text{Relevant\_Articles} \cap \text{top-n\_Articles\_Results}|}{n}$$

$$R@n = \frac{|\text{Relevant\_Articles} \cap \text{top-n\_Articles\_Results}|}{|\text{Relevant\_Articles}|}$$

$$F1 = 2\frac{P@n \cdot R@n}{P@n + R@n}.$$

Neither of P@n nor R@n alone is sufficient to assess the performance of any recommendation. This is the reason to introduce the F1 score. However, several points are to be noted here. First, in our context of recommendation, F1@n score is not as valuable as in the field of information retrieval for example. The recall R@n value is small when n is small and grows with n. Mathematically, it is not possible to capture a large proportion of relevant articles in the top-n results if the set of relevant results is much larger than n. So below $n = 10$, both recall R@n and F1@n are not very meaningful, which is the reason why our Table 3 gathers only these values for $n = 10$ and $n = 20$. In the latter case, the recall R@20 is quite good meaning that more than 70 % of the relevant articles are among the top-20 results. Second, we put ourselves in the context of a news Web platform that needs to recommend articles to its readers in order to hold them on the platform. In this case, our target is not to find the whole set of relevant articles, but rather a small set of articles for which we are sure that they are relevant. Hence, the relevant measure in our context is more the precision P@n for small values of n rather than the F1 score for high values of n. This is the reason for the focus on the precision in Table 4.

6.3 Results

The results of our experiments are presented in Tables 3 and 4. The former shows P@n, R@n and the corresponding F1 scores for $n = 10$ and $n = 20$ and for each recommendation strategy that we have described earlier. The latter shows the precision values P@n for $n \in \{1, 3, 5, 10, 20\}$ and for each recommendation strategy. We can clearly see that the Diversity strategy outperforms the baseline approaches by a significant margin: up to 17 % in terms of $P@5$ compared to NoEnrich and up to 21 % compared to AllComments. The F1 score at 20 also raises by 12 % from NoEnrich to Diversity, and even by 19 % from AllComments to Diversity. Having a closer look at the results, we notice that relying only on news

Table 3 Overall performance of our approach

	P@10	R@10	F1@10	P@20	R@20	F1@20
NoEnrich	0.513	0.361	0.423	0.540	0.715	0.615
AllComments	0.453	0.316	0.372	0.503	0.684	0.580
Authority	0.550	0.369	0.441	0.565	0.743	0.642
Diversity	0.640	0.425	0.511	0.607	0.810	0.694

Table 4 Overall precision of our approach

	P@1	P@3	P@5	P@10	P@20
NoEnrich	0.424	0.494	0.481	0.513	0.540
AllComments	0.393	0.474	0.445	0.453	0.503
Authority	0.454	0.535	0.530	0.550	0.565
Diversity	0.575	0.646	0.654	0.640	0.607

articles content (NoEnrich) does not give a good precision. Furthermore, even when trying to enrich news articles using all user comments (AllComments), both the precision and the F1 score decrease. By applying the ranking model and enriching news articles content by the most authoritative comments (Authority), the precision increases but the gain is small, ranging from 1 % to 4 % in Precision and about 4 % in F1 score. However, when we apply diversification to the most authoritative comments (Diversity), we obtain the best results.

We have tested the impact of changing k on both Authority and Diversity strategies. The results are shown in Table 5. First, we can observe that selecting the top-k authoritative comments to enrich news article contents performs better than using all comments. However, due to redundancies, this method becomes less effective especially when k increases, which is the case of $k = 20$. A similar observation holds for Diversity. We can see that increasing k from 5 to 10 improves the precision and actually provides the best results with respect to all baseline approaches. Increasing k above this value would cover more latent aspects however it would increase the risk to deviate from the topic resulting in lower precision values.

To better understand the behavior of each strategy, we have analysed the profiles of a set of news articles. Table 6 shows example aspects extracted using NoEnrich, AllComments, Authority, and Diversity. We observe that AllComments strategy leads to aspects which are either too broad or far remote from the main topic of the news article. Such remote aspects might be explained by the high quantity of noise inside the comments which deviates the news article profile from its main topic. These remote aspects damage the news article profile which is a key element for a further accurate recommendation. This explains that AllComments strategy provides a lower recommendation accuracy than NoEnrich strategy.

In the case of Diversity strategy, news article profiles are more relevant to the news article topic and contain more diverse aspects than other profiles. It is also clear that most of the aspects used to describe the news article are well linked to its topic. We can see in Table 6 that the aspects extracted for the news article "Donald Trump victory speech" are too generic with NoEnrich strategy. They express obvious aspects such as *Donald trump vote* and *Donald trump election*. On the other hand, we can see that Diversity strategy provides

Table 5 The impact of k on Authority and Diversity strategies

		P@1	P@3	P@5	P@10	P@20
	k=5	0.439	0.510	0.509	0.534	0.558
Authority	k=10	0.454	0.535	0.530	0.550	0.565
	k=20	0.439	0.530	0.521	0.553	0.559
	k=5	0.484	0.575	0.587	0.595	0.586
Diversity	k=10	0.575	0.646	0.654	0.640	0.607
	k=20	0.489	0.602	0.595	0.599	0.591

Table 6 News article profiles under the different strategies

Article Title	NoEnrich	AllComments	Authority	Diversity
British couple to be deported from Australia for living in wrong suburb	Australia Living, British Australia, couple Live, Australia life	Immigrants life, couple Law, Australian goverment, Australian suburb	Australian Visa, Australian Live, Australian Tax, Australian people, Australian government	Australian Tax, Australian people, Deportation, Australian Visa, Contract people, Live rules, living condition
Cameron needs to capture some of Boris Johnson sunshine	Boris Johnson, Cameron Johnson, Cameron party	Cameron party, Boris Johnson, Cameron voters, Election	Cameron policies, economy, Cameron finance policies, Economic leaders	Cameron Europe policy, Cameron finance policies, Business leader, Cameron party, Johnson policies, British economy
Argentina pulls out of Falklands talks	Island Falkland, Island Argentina, referendum	Falkland Argentina, Island referendum, Birtish colonialism, Falkland crisis	Island referundum, Argentina regime, British colonialism, Argentina Island	Falkland living, Falkland people, Colonialism, Referendum, Argentina government
Barack-Obama a dithering controlling risk averse	Obama policies, Clinton foreign policy, American policies	Obama foreign policy, president budget, world policy, Obama election	Island referundum, Argentina regime, British colonialism, Argentina Island	Afghanistan policy American foreign policies, Clinton policy, Foreign policy budget
Donald Trump victory speech	Donald Trump, Donald Trump election, Donald trump vote, Trump supporters	Donald Trump, Trump establishment, Good Trump, Crazy Trumpn, Trump favorites	Trump election, Trump vote, Clinton supporters, Trump president	Donald Trump supporters, Trump establishment, Clinton supporters, Trump election, , Trump international policy, Trump business, Donald Trump trickster

more and concise aspects such as *Trump establishment*, *Clinton supporters*, *Trump international policy*, and *Trump business*. A similar observation holds for the news article “British couple to be deported from Australia for living in wrong suburb”. The extracted aspects are too generic with NoEnrich and All Comments strategies. They are mainly about Australian life. By contrast, the aspects become more focused with comment ranking and deal for example with *Australian Visa* and *Deportation*. We can also notice that Diversity strategy extracts more aspects such as *Australia tax* and *people contracts*.

6.4 Discussion

The experiment results meet our expectations since they perfectly reflect the role and the nature of user comments in news websites. Relying only on articles content does not perform well because user profiles built on comments focus on some aspects that might differ from the ones provided by the news article. Figure 4 shows the proportions of aspects extracted from news articles and latent aspects extracted from user comments. With all comments, the description of the news article is mainly based on user comments. Clearly taking all user comments into account is not a good idea because they are subject to noise and some of them might even deviate from the topic of interest. Indeed, this approach has achieved the lowest rate of performance.

Selecting the top-k user comments reduces the number of covered latent aspects but the recommendation quality does not improve significantly. This can be explained by redundancies that appear in the most authoritative user comments. It is very common for users to express their opinions in debates several times trying to convince others. They can change arguments but they still talk about the same aspects. Since our approach focuses only on aspects, it can select different user comments on the topk but then extract from them exactly the same aspects. For this reason, we have adopted diversification.

By using diversification, we can cover more latent aspects without biasing the content of news articles which explains its good performance. The reason is the broader coverage of aspects. If the aspects discussed in the comments are not explicit in the news article, they are added to the ness article profile which increases the chance of having more users getting interested in the article. It is also important to note that covering 72 % of latent aspects with only the best 10 diversified comments is still large which might involve noise. This explains precision values not getting higher than 65.4 %.

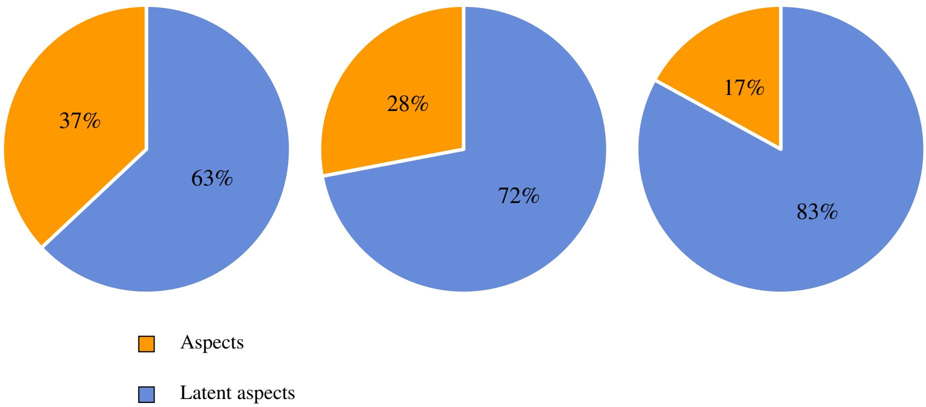

Figure 4 Proportions of aspects and latent aspects in news articles profiles

7 Conclusion

In this article, we addressed the problem of recommendation in the context of news Web sites. In particular, we used different ways to leverage user generated content on news articles for a more effective recommendation system. We proposed two approaches for enriching news article contents by user comments. The first approach employs only relevant comments using a ranking strategy, while the second approach uses diversified comments. Our study conducted on an extensive set of experiments shows that diverse comments achieve much better results compared to baseline approaches. It proves that user generated content is a valuable source of information that can capture user interests and contribute to the extension of knowledge about daily life topics.

As a future work, we aim at exploring the impact of co-comments patterns. To achieve this goal, we plan to extend our model to a hybrid recommender model in which we use collaborative filtering recommendation techniques. Another important point to address is to investigate user interests evolution over time. This could give useful insights on exploring user commenting and reading behaviors.

References

1. Abbar, S., Amer-Yahia, S., Indyk, P., Mahabadi, S.: Real-time recommendation of diverse related articles. In: Proceedings of the 22nd international conference on world wide Web, pp. 1–12. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva (2013)
2. Abel, F., Gao, Q., Houben, G.-J., Tao, K.: Analyzing user modeling on twitter for personalized news recommendations. In: Proceedings of the 19th international conference on user modeling, adaption, and personalization, pp. 1–12. Springer, Berlin (2011)
3. Aker, A., Kurtic, E., Balamurali, A., Paramita, M., Barker, E., Hepple, M., Gaizauskas, R.: A graph-based approach to topic clustering for online comments to news. In: Advances in information retrieval, pp. 15–29. Springer (2016)
4. Bansal, T., Das, M., Bhattacharyya, C.: Content driven user profiling for commentworthy recommendations of news and blog articles. In: Proceedings of the 9th acm conference on recommender systems, pp. 195–202. ACM, New York (2015)
5. Chen, J., Nairn, R., Nelson, L., Bernstein, M., Chi, E.: Short and tweet: Experiments on recommending content from information streams. In: Proceedings of the sigchi conference on human factors in computing systems, pp. 1185–1194. ACM, New York (2010)
6. Danescu-Niculescu-Mizil, C., Kossinets, G., Kleinberg, J., Lee, L.: How opinions are received by online communities: a case study on amazon.com helpfulness votes. In: Proceedings of the 18th international conference on world wide Web, pp. 141–150. ACM, New York (2009)
7. Ganesan, K., Zhai, C.: Opinion-based entity ranking. Inf. Retr. **15**(2), 116–150 (2012)
8. Gollapudi, S., Sharma, A.: An axiomatic approach for result diversification. In: Proceedings of the 18th international conference on world wide Web, pp. 381–390. ACM, New York (2009a)
9. Gollapudi, S., Sharma, A.: An axiomatic approach for result diversification. In: Proceedings of the 18th international conference on world wide Web, pp. 381–390. ACM, New York (2009b)
10. Hassin, R., Rubinstein, S., Tamir, A.: Approximation algorithms for maximum dispersion. Oper. Res. Lett. **21**, 133–137 (1997)
11. Hu, M., Sun, A., Lim, E.-P.: Comments-oriented document summarization: Understanding documents with readers' feedback. In: Proceedings of the 31st annual international acm sigir conference on research and development in information retrieval, pp. 291–298. ACM, New York (2008)
12. Kant, R., Sengamedu, S.H., Kumar, K.S.: Comment spam detection by sequence mining. In: Proceedings of the fifth acm international conference on Web search and data mining, pp. 183–192. ACM, New York (2012)
13. Kim, S., Pantel, P., Chklovski, T., Pennacchiotti, M.: Automatically assessing review helpfulness. In: Proceedings of the 2006 conference on empirical methods in natural language processing, pp. 423–430 (2006)
14. Korte, B., Hausmann, D.: An analysis of the greedy heuristic for independence systems. Ann. Discret. Math. **2**, 65–74 (1978)

15. Li, Q., Wang, J., Chen, Y.P., Lin, Z.: User comments for news recommendation in forum-based social media. Inform. Sci. **180**(24), 4929–4939 (2010a)
16. Li, Q., Wang, J., Chen, Y.P., Lin, Z.: User comments for news recommendation in forum-based social media. Inf. Sci **180**(24), 4929–4939 (2010b)
17. Lin, C., He, Y.: Joint sentiment/topic model for sentiment analysis. In: Proceedings of the 18th acm conference on information and knowledge management, pp. 375–384. ACM, New York (2009)
18. Litvak, M., Matz, L.: Smartnews: Bringing order into comments chaos. In: Proceedings of the international conference on knowledge discovery and information retrieval, kdir, vol. 13 (2013)
19. Liu, C.-Y., Chen, M.-S., Tseng, C.-Y.: Incrests: Towards real-time incremental short text summarization on comment streams from social network services. IEEE Trans. Knowl. Data Eng. **27**, 2986–3000 (2015)
20. Liu, Y., Huang, X., An, A., Yu, X.: Modeling and predicting the helpfulness of online reviews. In: Eighth ieee international conference on data mining, 2008. icdm'08, pp. 443–452 (2008)
21. Llewellyn, C., Grover, C., Oberlander, J.: Summarizing newspaper comments. In: Icwsm (2014)
22. Marcheggiani, D., Täckström, O., Esuli, A., Sebastiani, F.: Hierarchical multilabel conditional random fields for aspect-oriented opinion mining. In: Advances in information retrieval, pp. 273–285. Springer (2014)
23. Meguebli, Y., Kacimi, M., Doan, B.-L., Popineau, F.: Unsupervised approach for identifying users? political orientations. In: Advances in information retrieval, pp. 507–512. Springer (2014)
24. Persing, I., Ng, V.: Vote prediction on comments in social polls. In: Emnlp, pp. 1127–1138 (2014)
25. Phelan, O., McCarthy, K., Smyth, B.: Using twitter to recommend real-time topical news. In: Proceedings of the third acm conference on recommender systems, pp. 385–388. ACM, New York (2009)
26. Popescu, A.-M., Etzioni, O.: Extracting product features and opinions from reviews. In: Proceedings of the conference on human language technology and empirical methods in natural language processing, pp. 339–346. Association for Computational Linguistics, Stroudsburg (2005)
27. Rayana, S., Akoglu, L.: Collective opinion spam detection: Bridging review networks and metadata. In: Proceedings of the 21th acm sigkdd international conference on knowledge discovery and data mining, pp. 985–994. ACM, New York (2015)
28. Rendle, S., Freudenthaler, C., Gantner, Z., Schmidt-Thieme, L.: Bpr: Bayesian personalized ranking from implicit feedback. In: Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence, pp. 452–461. AUAI Press, Arlington (2009)
29. San Pedro, J., Yeh, T., Oliver, N.: Leveraging user comments for aesthetic aware image search reranking. In: Proceedings of the 21st international conference on world wide Web, pp. 439–448. ACM, New York (2012)
30. Shmueli, E., Kagian, A., Koren, Y., Lempel, R.: Care to comment?: Recommendations for commenting on news stories. In: Proceedings of the 21st international conference on world wide Web, pp. 429–438. ACM, New York (2012)
31. Terra, E., Clarke, C.L.A.: Frequency estimates for statistical word similarity measures. In: Proceedings of the 2003 conference of the north american chapter of the association for computational linguistics on human language technology, vol. 1, pp. 165–172. Association for Computational Linguistics, Stroudsburg (2003)
32. Tsagkias, M., Weerkamp, W., de Rijke, M.: Predicting the volume of comments on online news stories. In: Proceedings of the 18th acm conference on information and knowledge management, pp. 1765–1768. ACM, New York (2009)
33. Tsagkias, M., Weerkamp, W., de Rijke, M.: News comments: Exploring, modeling, and online prediction. In: Proceedings of the 32nd european conference on advances in information retrieval, pp. 191–203. Springer, Berlin (2010)
34. Tsur, O., Rappoport, A.: Revrank: a fully unsupervised algorithm for selecting the most helpful book reviews. In: International aaai conference on weblogs and social media (2009)
35. Wang, H., Lu, Y., Zhai, C.: Latent aspect rating analysis on review text data: a rating regression approach. In: Proceedings of the 16th acm sigkdd international conference on knowledge discovery and data mining, pp. 783–792. ACM, New York (2010)
36. Ye, J., Kumar, S., Akoglu, L.: Temporal opinion spam detection by multivariate indicative signals (2016). arXiv:1603.01929
37. Yee, W.G., Yates, A., Liu, S., Frieder, O.: Are Web user comments useful for search. In: Proceedings of LSDS-IR, pp. 63–70 (2009)
38. Zhao, L., Huang, M., Sun, J., Luo, H., Yang, X., Zhu, X.: Sentiment extraction by leveraging aspect-opinion association structure. In: Proceedings of the 24th acm international on conference on information and knowledge management, pp. 343–352. ACM, New York (2015)
39. Zhuang, L., Jing, F., Zhu, X.-Y.: Movie review mining and summarization. In: Proceedings of the 15th acm international conference on information and knowledge management, pp. 43–50. ACM, New York (2006)