Information Sciences 180 (2010) 4929–4939

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

User comments for news recommendation in forum-based social media

Qing Li [a,*], Jia Wang [a], Yuanzhu Peter Chen [b], Zhangxi Lin [c]

[a] Southwestern University of Finance and Economics, China
[b] Memorial University of Newfoundland, Canada
[c] Texas Tech University, USA

ARTICLE INFO

Article history:
Received 1 December 2009
Received in revised form 20 August 2010
Accepted 26 August 2010

Keywords:
News recommendation
Recommender system
Content-based filtering
Collaborative filtering
Social media
User comment
Information retrieval

ABSTRACT

News recommendation and user interaction are important features in many Web-based news services. The former helps users identify the most relevant news for further information. The latter enables collaborated information sharing among users with their comments following news postings. This research is intended to marry these two features together for an adaptive recommender system that utilizes reader comments to refine the recommendation of news in accordance with the evolving topic. This then turns the traditional "push-data" type of news recommendation to "discussion" moderator that can intelligently assist online forums. In addition, to alleviate the problem of recommending essentially identical articles, the relationship (*duplicate*, *generalization*, or *specialization*) between recommended news articles and the original posting is investigated. Our experiments indicate that our proposed solutions provide an improved news recommendation service in forum-based social media.

1. Introduction

In its brief history of about two decades, the Web has evolved from a technical framework for information dissemination to more of an enabler of social interactions among its users. During the earlier days, the Web was dominated by static pages or sites. Later on, dynamic content generation was added as enhancement of Web server intelligence. At the same time, Web browsers providing client-side calculation and event handling, Web applications became a prevalent framework for distributed GUI applications. Such technological advancement has fertilized vibrant creation, sharing, and collaboration among Web users. As a result, the role of Computer Science is not as much of designing or implementing certain data communication techniques, but more of facilitating a variety of creative uses of the Web [1].

In a more general context, nowadays, Web is one of the most important vehicles for "social media". Examples of social media include Internet forums, blogs, wikis, podcasts, instant messaging, social networking etc. One form of social media of particular interest here is self-publication, or user-generated media. In self-publication, a user can publish an article or post news to share with others. Other users can read and comment on the posting and these comments can, in turn, be read and commented on. Digg (www.digg.com) and Yahoo!Buzz (buzz.yahoo.com) are commercial examples of self-publication. Apparently, this is an extension of Internet forums with some quantitative metrics. A useful extension of this forum application is to add a recommendation feature to the current discussion thread. That is, based on the (1) original posting, (2) various levels of comments, and (3) their votes, the system can provide a set of relevant articles, which are expected to be of interest of the active users of the thread. The users learning experiences with the system can be immensely enhanced

* Corresponding author. Tel./fax: +86 28 87356198.
E-mail address: kooliqing@gmail.com (Q. Li).
URL: http://fife.swufe.edu.cn/BILab/ (Q. Li).

0020-0255/$ - see front matter
doi:10.1016/j.ins.2010.08.044

with the recommended articles. Due to the evolving nature of the thread and the inherent small sizes of the comments, such personalized recommendation will be different from traditional news recommendation to a large extent. Indeed, the evolution of discussion threads has been observed in USENET newsgroups [23].

In this article, we study the problem of news recommendation for dynamic discussion threads as an aspect of social media adaptation. In particular, we focus on the analysis of the news entity and the reader comment contents to improve the performance of the adaptive news recommendation.

2. Motivation and contribution

A fundamental challenge in adaptive news recommendation is to account for topic divergence, i.e. the change of gist during the process of discussion. In a forum, the original news is typically followed by other readers' opinions, in the form of comments and votes. Concerns and intention of active users may change as the discussion continues. Therefore, news recommendation, if it were only based on the original posting, can hardly benefit the potentially changing interests of the users. For instance, suppose that the original news posting is on deforestation, but the discussion has later drifted to the topics of desertification and global warming. Apparently, there is a need to consider topic evolution in adaptive news recommendation and this entails novel techniques that can help to capture topic evolution precisely.

The topic evolution of news recommendation in social media can be reflected by the comments. However, the impact of each comment varies according to its quality. In particular, some comments reflect insightful opinions on news stories, while others can be completely meaningless. Distinguishing the importance of each comment is crucial to utilize them properly in guiding topic evolution for recommending news in social media. Such differentiation can be simply done by considering user votes for each comment in social media. However, we argue that the content and structural linkage (e.g. quotation and reply) of comments should also be utilized to measure the quality of comments. This information is especially helpful to measure the quality of comments in case of lacking user votes or vote abuse.

Traditional news recommendation systems focus on selecting the most interesting articles to the users. However, readers can be burdened by the more or less identical ones among the returned news articles. An interpretation of the connection between recommended news and the original posting would be useful to assist users in further reading and helping readers skip duplicate articles. In particular, readers would usually appreciate an article being labeled as generalization, specialization, or duplicate of the original posting. Indeed, the value of interpreting recommendation results in enhancing users' trusting beliefs has been observed by Wei and Benbasat [29]. Apparently, the challenge lies in how to generate the interpretation.

In this work, we propose a framework for adaptive news recommendation in social media. It has the following unique features that have not been attempted in previous work:

- It is part of our first attempt of incorporating reader comments for adaptive news recommendation so that relevant information is recommended based on a balanced perspective of both the authors and readers. To do that, we model the relationship among comments and that relative to the original posting in order to evaluate their overall effect.
- We determine the relationship (*duplicate*, *generalization* or *specialization*) between suggested news articles and the original posting, and present it to users via a novel user interface.

The rest of this work is organized as follows. We first briefly describe necessary background techniques in Section 3. The design details for this news recommendation framework are presented in Section 4. We then test the performance of such a recommender using a news corpus and discussion threads that have been collected by a crawler (Section 5). This paper is concluded with speculation on how the current prototype can be further improved in Section 6.

3. Related work

In a broader context, a related problem is content-based information recommendation (or filtering). Most information recommendation systems select articles based on the contents of the original postings. The relevant information selections of these work are determined by the textual similarity between the recommended news and the original news posting. For instance, Chiang and Chen [7] study a few classifiers for agent-based news recommendations. Zheng et al. [33] exploit semantic relationships among nouns and noun phrases for document retrieval.

A number of later proposals incorporate additional metadata, such as user behaviors, domain knowledge, and timestamps. For example, Guo et al. [11] propose cognitive situation models to exploit a human cognitive procedure in categorizing texts. Liu et al. [22] utilize sequences of user mouse clicking to suggest relevant news. Su et al. [28] recommend products based on the profiles built on the product contents and user behaviors. Claypool et al. [8] recommend breaking news with numerical user ratings. Lee and Park [17] consider matching between news article attributes and user preferences. Anh et al. [1] and Lai et al. [14] construct explicit user profiles, respectively. Lavrenko et al. [16] propose the e-Analyst system which combines news stories with financial knowledge including stock price trends. Cantador et al. [6] utilize domain ontology to recommend news. Del Corso et al. [10] use timestamps to favor more recent news. Some go even further by ignoring the news contents and only using the browsing behaviors of the readers with similar interests [9,18,20].

Another related problem is topic detection and tracking (TDT), i.e. automated categorization of news stories by their themes. TDT consists of breaking the stream of news into individual news stories, monitoring the stories for events that have not been seen before, and categorizing them [15]. A topic is modeled with a language profile deduced by the news. Most existing TDT schemes calculate the similarity between a piece of news and a topic profile to determine its topic relevance [15,30,31]. Qiu et al. [25] apply TDT techniques to group news for collaborative news recommendation. Some work on TDT takes one step further in that they update the topic profiles as part of the learning process during its operation [2,19]. Note that most existing work emphasizes on the precision of selecting relevant news for recommendation. Few study the novelty of recommendation, i.e. the ability of recommending nonidentical relevant news.

In summary, in contrast to news recommendation in traditional media, which recommends news based on the semantic similarity between the original news posting and incoming news articles [21], we argue that, in social media, especially news discussion forums, more effective news recommendation should consider the very important fact of the evolution of discussion thread. Meanwhile, a good interpretation of the relationship between recommended news and the original posting will alleviate duplicate recommendation.

4. Framework of adaptive news recommendation

In this section, we present a mechanism for adaptive news recommendation. The framework is sketched in Fig. 1. Essentially, it builds a topic profile for each original news posting along with the comments from readers, and uses this profile to retrieve relevant news. In particular, we first model the relationship among comments and that relative to the original posting in order to evaluate their overall impact. This information along with the news and comments is fed into a synthesizer. The synthesizer balances views of both authors and readers to construct a topic profile to retrieve relevant news. These recommended news articles are further analyzed by an interpreter to reveal their relationship to the original posting.

4.1. Scoring comments

In social media, comments reveal readers' concerns about a more general or specific topic of the original news posting to some degree. However, the impact of each comment varies according to its quality. In this section, we use a graph-based model to describe the connections of comments, and utilize this model to score comments.

In a discussion thread, comments made at different levels reflect the variation of focus of its participants (Fig. 2). Therefore, comments should be incorporated when building recommendation models. In our model, we treat the original posting and the comments each as a text node. This model considers both the content similarity between text nodes and the logic relationship among them.

On one hand, the content similarity between two nodes can be measured by any commonly adopted metric, such as cosine similarity and Jaccard coefficient. This metric is taken over every node pair in the discussion thread. On the other hand, the logic relation between nodes takes two forms. First, a comment is always made in response to the original posting or an earlier comment. In graph theoretic terms, the hierarchy can be represented as a tree $T = (V, E_T)$, where V is the set of all text nodes and E_T is the set of edges. In particular, the original posting is the root and all the comments are ordinary nodes. There is a directed edge $e \in E_T$ from node u to node v, denoted (u, v), if the corresponding comment v was made in response to comment (or original posting) u. Second, a comment can quote from one or more earlier comments. From this perspective, the hierarchy can be modeled using a directed acyclic graph (DAG), denoted $D = (V, E_D)$. There is a directed edge $e \in E_D$ from node u to node v, denoted (u, v), if the corresponding comment v quotes from comment (or original posting) u. As shown in Fig. 3, for either graph T or D, we can use a $|V| \times |V|$ adjacency matrix, denoted M_T and M_D, respectively, to record them. Inline with the adjacency matrices, we can also use a $|V| \times |V|$ matrix defined on $[0, 1]$ to record the content similarity between nodes and denote it by M_C. Thus, we can combine these three aspects linearly:

$$M = \gamma_1 \times M_C + \gamma_2 \times M_T + \gamma_3 \times M_D,$$

where M is a $|V| \times |V|$ adjacency matrix capturing both semantic and structural relation among text nodes.

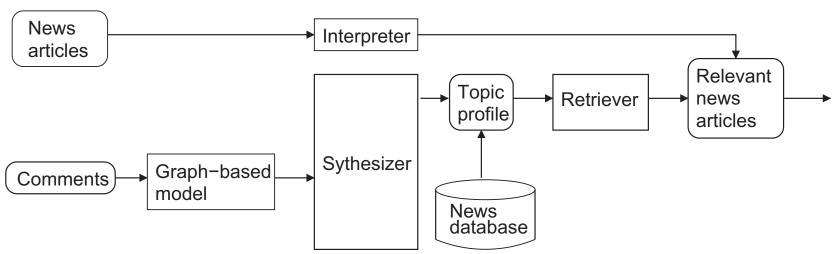

Fig. 1. Design scheme.

Fig. 2. A news article chain of comments.

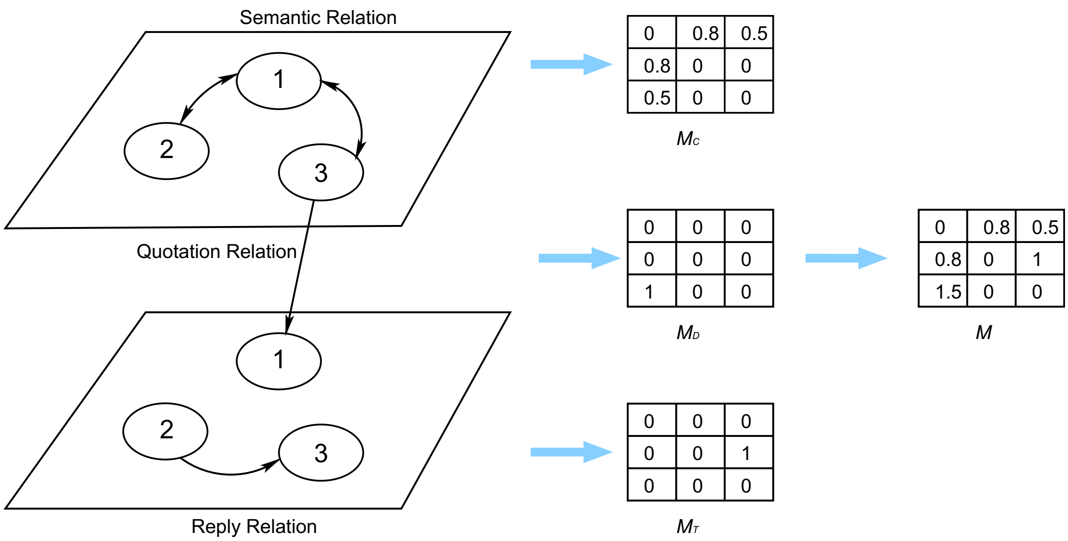

Fig. 3. Multi-relation graph of comments.

Intuitively, the important comments are those whose topics are discussed by a large number of other important comments. Therefore, we use a variant of the PageRank algorithm [5] to rank the comments as

$$s_j = \frac{\lambda}{|V|} + (1-\lambda) \times \sum_{c_i} r(c_i, c_j) \times s_i,$$

where λ is the damping factor as in PageRank and its value is recommended to be 0.85, i and j are node indices, and $|V|$ denotes the number of text nodes in the thread. In addition, $r(c_i, c_j)$ is the normalized weight of comment c_i referring to c_j defined as

$$r(c_i, c_j) = \frac{M_{i,j}}{\sum_{c_j} M_{i,j} + \epsilon},$$

where $M_{i,j}$ is an entry in the graph adjacency matrix and ϵ is a constant to avoid division by zero.

4.2. Topic profile construction

Once the importance of comments on one news posting is quantified by our graph-based model, this information along with the news itself is fed into a synthesizer to construct a topic profile of this news discussion thread. As such, the views from the authors and readers are balanced. Note that, in some systems, votes for comments are available to provide a metric to quantify comments. Here, we leave this for future investigation for two reasons. First, without relying on votes, our framework is more universal in terms of usefulness as many social media do not support voting yet. Second, at this point, most Web sites do not have effective measures against user tampering votes.

Our implementation of the profile is a weight vector of all terms to model the language used in the thread. The vocabulary here is the words in the English language. Consider a news posting $d_0$and its comment sequence $\{d_1,d_2,\ldots,d_m\}$. For each term t, a compound weight $W(t) = \alpha \times W_1(t) + (1-\alpha) \times W_2(t)$ is calculated. It is a linear combination of the contribution by the news posting itself, $W_1(t)$, and that by the comments, $W_2(t)$. We assume that each term is associated with an "inverted document frequency" with regard to the training news corpus, denoted by $I(t) = \log\frac{N}{n(t)}$, where N is the corpus size and $n(t)$ is the number of documents in corpus containing term t. As convention, we use a binary function $f(t,d)$ to denote the number of occurrences of term t in document d, i.e. "term frequency". Thus, when news posting and comments are each considered as a document, this term frequency value can be calculated for any term in any document. We thus define the weight of term t in document d as below, be it the news posting itself or a comment, using the standard TF $\times$ IDF definition [3]:

$$w(t,d) = \left(0.5 + 0.5 \times \frac{f(t,d)}{\max_{t'} f(t',d)}\right) \times I(t).$$

The weight contributed by the news itself, d_0, is thus:

$$W_1(t) = \frac{w(t,d_0)}{\max_{t'} w(t',d_0)}.$$

The weight contribution from the comments $\{d_1,d_2,\ldots,d_m\}$ incorporates not only the language features of these documents but also their importance of leading a discussion in related topics. That is, the contribution of comment score is incorporated into weight calculation of the words in a text node:

$$W_2(t) = \sum_{i=1}^{m} \frac{w(t,d_i)}{\max_{t'} w(t',d_i)} \times \frac{s_i}{\max_{i'} s_{i'}}.$$

Such a treatment of compounded weight $W(t)$ is essentially to recognize that readers' impact on selecting relevant news and the difference of their influences. For each profile, we select the most weighted words to represent the topic.

4.3. News retrieval with relevance language models

With the topic profile constructed as above, we can use it to select relevant news from a collection of news articles in the database. That is, the retriever returns an ordered list of news articles with decreasing relevance to the topic. Our graph model can be incorporated into any good news retrieval model to differentiate the importance of each comment. In this work, we combine the strengths of two state-of-the-art news retrievers (time factor [10] and language model [16]) to construct a powerful retrieval module.

Assume that each news posting d is associated with the time when posted. Upon recommendation, our retrieval algorithm considers two factors for recommendation candidates: (1) the relevance based on the profile similarity, and (2) the timeliness based on the time stamp. Specifically, the score of the candidate document d with regard to profile p is defined as

$$s(p,d) = (p \cdot d) \times e^{\tau/\beta},$$

where the first factor is the similarity of the language vectors which can be measured by any commonly adopted metric, such as cosine similarity and Kullback–Leibler (KL) divergence, τ is the number of days that has elapsed since d was posted, and β is a constant to tune the scale of time attenuation, set to 365 to simulate an annual attenuation rate.

In this work, to measure the content similarity $(p \cdot d)$, we take the relevance language model approach of Lavrenko and Croft [15]. The similarity between a news article and a topic profile is measured by the KL divergence between the document model and the topic model. Articles with smaller divergence are considered more relevant. The divergence is defined as:

$$(p \cdot d) \approx KL(M_p \| M_d) = \sum_{w} P(w|M_p) \log\frac{P(w|M_p)}{P(w|M_d)},$$

where w denotes a word, M_d is a language model for document d in the collection, which is a probability distribution that captures the statistical regularities of the natural language. M_p is a language model for topic p. $P(w|M_p)$ can be estimated as:

$$P(w|M_p) \approx \frac{P(w,m_1\cdots m_k)}{P}(m_1\cdots m_k) = \frac{P(w,m_1\cdots m_k)}{\sum_t P(t,m_1\cdots m_k)},$$

$$P(w,m_1\cdots m_k) = \sum_{d\in D_r} P(M_d)\left[P(w|M_d)\prod_{i=1}^{k} P(m_i|M_d)\right],$$

where m_i denotes the ith word of the topic profile, t denotes a word in the vocabulary. Here, $P(M_d)$ denotes *a prior* distribution over the relevant document set D_r, usually uniform. To expedite execution, we restrict ourselves to only 50 top-ranked documents retrieved by the topic profile. $P(w|M_d)$ specifies the probability of observing w if we take a word randomly from M_d. It is estimated by

$$P(w|M_d) = \eta \times \frac{f_{w,d}}{|d|} + (1-\eta) \times \frac{f_w}{F},$$

where $f_{w,d}$ is the number of times word w occurs in document d, f_w is the number of times word w occurs in the corpus. F is the total number of tokens (i.e. words with repetition) in the corpus, $|d|$ is the number of tokens in d. Note that we use a parameter η to control the contribution of the term frequency to this probability. $P(m_i|M_d)$ is estimated in the same way.

With such a scoring mechanism, the retriever can return a set of most relevant, timely news articles with regard to any evolving discussion in a self-publication system.

4.4. Recommendation interpretation

We next present a labeling scheme to categorize a recommended article as specialization, generalization, and duplicate relative to the original posted news article. Since a recommended article is always relevant to the original, the diversity of words used in an article is a good indicator of the breadth of the scope. Such an indicator is calculated using the information entropy carried by the article.

We consider a corpus and denote its vocabulary as a set of n words $\{w_1, w_2, \ldots, w_n\}$. Given an article A in the corpus, we can calculate the TF $\times$ IDF for the words herein and normalize it as a vector $\vec{A} = \langle a_1, a_2, \ldots, a_n \rangle$, where $\sum_{i=1}^{n} a_i = 1$. This vector can be treated as a probability distribution function of the vocabulary in article A. Thus, the entropy $H(A)$ reflects the generality and speciality of the article, where

$$H(\vec{A}) = \sum_{i=1}^{n} a_i \times \log \frac{1}{a_i}.$$

Apparently, the less the entropy is, the more specialized an article is. Consider another article B in the corpus. We define the vector for article B similarly as $\vec{B} = \langle b_1, b_2, \ldots, b_n \rangle$. Assuming that articles A and B are relevant, this allows us to define their relative relationship (generalization, specialization, and duplicate) by comparing their entropies $H(\vec{A})$ and $H(\vec{B})$. (Note that when A is a generalization of B, B is a specialization of A.)

In order to accommodate the situation when A and B are considered duplicate of each other, we define a case where their entropies are "sufficiently close" as follows. We denote the set of common words that appear in both A and B by $\mathcal{C}$. For article A, the entropy $H(\vec{A})$ is contributed by the words in $\mathcal{C}$, denoted by $H_A^{\mathcal{C}}$, and by the remaining words, denoted by $H_A^{\bar{\mathcal{C}}}$. That is,

$$H(\vec{A}) = H_A^{\mathcal{C}} + H_A^{\bar{\mathcal{C}}}.$$

Similarly, the entropy for article B can be decomposed as

$$H(\vec{B}) = H_B^{\mathcal{C}} + H_B^{\bar{\mathcal{C}}}.$$

Such relations are visually represented in Fig. 4.

When the common part of articles A and B contributes a significant portion to the total entropies, we say that A and B are duplicate of each other, i.e.

$$\frac{H_A^{\mathcal{C}}}{H(\vec{A})} \geqslant \eta \quad \text{or} \quad \frac{H_B^{\mathcal{C}}}{H(\vec{B})} \geqslant \eta,$$

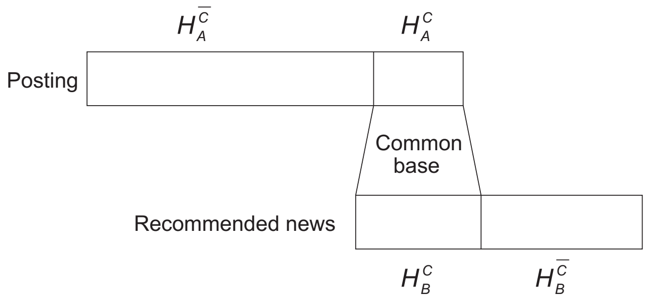

Fig. 4. Visual representation of the contents of the original posting A and recommended news B.

where η is set to 0.7 based on our preliminary tests. In our preliminary experiments, we find that neither too small nor too large a value is appropriate for interpretation. When neither condition holds, we compare $H_A^{\bar{C}}$ with $H_B^{\bar{C}}$. Recall that the complementary portion's entropy is a metric for the degree of semantic extension over the common part. When $H_A^{\bar{C}} > H_B^{\bar{C}}$, we say that *A* is a generalization of *B*; otherwise, we say that *A* is a specialization of *B*.

5. Experimental evaluation

In this section, we present the experiments to evaluate our proposed approach. We first study the overall performance of the recommendation system. Then we investigate the effect of integrating reader comments semantically and structurally. Last, we observe how recommendation interpretation alters the system behaviors.

5.1. Experimental settings

To gauge how well our recommender performs, we carry out our experiments on a synthetic data set collected from the Digg Web site (www.digg.com) and Reuters News (www.reuters.com). We randomly select 45 news articles along with reader comments from Digg, each of which is called a *topic* here. These topics are treated as the originals, and recommended news are selected from a corpus of articles collected from Reuters. This simulates the scenario of recommending relevant news from traditional Web sites to social networking Web site users for their further reading. For evaluation purposes, we adopt the traditional pooling strategy [32] applied to the TREC data set[1] at the moment to mark the relevant articles for each topic. In particular, we use three different retrieval engines (i.e. Lucence, Indri, and Okapi) to produce a list of potentially relevant news articles. To do that, each retrieval engine is queried separately first. Then we merge their returned articles to generate a master ranking. Furthermore, we label each article from the master ranking with generalization, specialization, or duplicate relative to the topic. A summary of our data set is shown in Table 1.

5.2. Baseline methods

In addition to our proposed news recommender, we also implement two well-known news recommendation methods as the baseline [4].

The first method, Okapi, is commonly applied as a representative of the classic probabilistic model, i.e., 2-poisson model, for relevant information retrieval [26]. The second one, LM, is based on statistical language models for relevant information retrieval. It builds a probabilistic language model for each article, and ranks them on query likelihood, i.e. the probability of the model generating the query [24]. Following the strategy of Bogers and Bosch [4], relevant articles are selected based on the title and the first 10 sentences of the original postings. This is because articles are organized in a so-called *inverted pyramid* style, meaning that the most important information is usually placed at the beginning. Trimming the rest of an article would usually remove relatively less crucial information, which speeds up the recommendation process.

5.3. Metrics

Traditionally, the performance of news recommendation schemes is evaluated by the precision metric. For news recommendation, duplicate news can greatly reduce the quality of the recommender since few users would like to read the same articles repeatedly. Therefore, we argue that the innovativeness in the recommended articles is as important for effectiveness evaluation as precision itself. Each recommended article is labeled as specialization, generalization, or duplicate. Thus, we define the precision and innovativeness metrics as

$$P@N = \frac{|C \cap R|}{|R|}, \quad I@N = \frac{|E \cap R|}{|R|},$$

where *R* is the subset of the top-*n* recommended articles, *C* is the set of relevant articles, and *E* is the above set with duplicate removed. We select top-10 articles for evaluation assuming most of readers only browse up to 10 items down the list [13]. Meanwhile, we also take the mean average precision (MAP) and mean average innovativeness (MAI) as metrics to evaluate the entire set of returned articles.

5.4. Statistical significance

Before reporting our experimental results using precision and innovativeness, we provide a significance test to show that the observed differences are not incidental.

The Wilcoxon signed-rank test and *t*-test are commonly used for the significance test in information retrieval experiments [27]. In a nutshell, both take a pair of equal-sized sets of per-query effectiveness values, and assign a confidence value

[1] A standard textual data set used in information retrieval research posted at http://trec.nist.gov/data.html.

Table 1
Data set summary.

Number of topics	45
Average length (word count) in original news articles	685
Average number of comments in each topic	87
Average length (word count) of comments	53
Number of news articles in corpus	25,650
Average length (word count) of news articles in corpus	569

to the null hypothesis that the values are drawn from the same distribution. If confidence in the hypothesis (reported as a *p*-value) is less than 5%, it typically means the results of experiments are reliable and convincible.

Among the assumptions of the Wilcoxon signed-rank test and the *t*-test are that the values being tested – in our case, per-query effectiveness – are distributed symmetrically and normally, respectively [27]. However, effectiveness rarely follows either distribution. Instead, Hull [12] points out that the *t*-test can be reliable even when data being tested are not distributed normally. Therefore, we applied a paired *t*-test to find out whether the observed difference is incidental.

5.5. Overall performance

A paired *t*-test shows that using $P@10$ and $I@10$ as performance measures, our approach performs significantly better than the baseline methods as shown in Table 2, at $p = 0.012$ and $p = 0.025$, respectively. In addition, we have *t*-tests using MAP and MAI as performance measures, respectively, and the *p* values of these tests are all less than 0.05, which means that the results of experiments are statistically significant. We believe that such gains are introduced by the additional information from the comments.

5.6. Parameters of topic profile

There are two important parameters (Section 4.2) to be considered when constructing topic profiles for recommendation. (1) The number of the most weighted words to represent the topic, and (2) combination coefficient α to determine the contribution of original posting and comments in selecting relevant articles. We conduct a series of experiments and find out that the optimal performance is obtained when the number of words is between 50 and 70, and α is between 0.65 and 0.75. When α is set to 1, the recommended articles only reflect the author's opinion. When $\alpha = 0$, the suggested articles represent the concerns of readers exclusively. In the following experiments, we set topic word number to 60 and combination coefficient α to 0.7.

5.7. Effect of comments

To observe the impact of readers' concerns about original news posting in social media, which is reflected by comments on the news postings, we study the following scenarios:

- RUN1 (News): the topic file is constructed only based on the original news posting itself. This is analogous to the traditional news recommenders which only consider the focus of authors for suggesting further readings.
- RUN2 (Comments): the topic file is constructed only based on the comments.
- RUN3 (Both): the topic file is constructed based on both the content of news and its comments.

Here, we set $\gamma_1 = \gamma_2 = \gamma_3 = 1$. Our *t*-test shows that using $P@10$ and $I@10$ as performance measures, RUN3 performs best as shown in Table 3, at $p = 0.021$ and $p = 0.014$, respectively. These are much less than the critical confidence value (0.05). We believe that such gains are contributed by the additional information from user comments.

Furthermore, we investigate the effect of the three forms of relationships among comments, i.e. content similarity, reply, and quotation. We carry out a series of experiments for this purpose. Recall that a graph-based model integrating content and physical link relation is applied to distinguish comments (Section 4.1). Therefore, we adjust the settings of this graph-based model to investigate the power of these factors.

Table 2
Overall performance.

Method	Precision		Innovativeness	
	$P@10$	MAP	$D@10$	MAI
Okapi	0.838	0.83	0.812	0.78
LM	0.834	0.813	0.815	0.785
Our work	0.947	0.938	0.91	0.86

Table 3
Performance of three runs.

Method	Precision		Innovativeness	
	P@10	*MAP*	*I@10*	*MAI*
RUN1 (news)	0.887	0.87	0.858	0.8
RUN2 (comments)	0.923	0.9	0.902	0.853
RUN3 (both)	0.947	0.938	0.91	0.86

- Content relation (CR): only the content relation matrix is used in scoring the comments instead of the multi-relation affinity matrix.
- Quotation relation (QR): only the quotation relation matrix is used in scoring the comments.
- Reply relation (RR): only the reply relation matrix is used in scoring the comments.
- Content + quotation relation (CQR): both the content and quotation relation matrices are used in scoring the comments.
- Content + reply relation (CRR): both the content and reply relation matrices are used in scoring the comments.
- Quotation + reply relation (QRR): both the quotation and reply relation matrices are used in scoring the comments.
- All: all three matrices are used.

As shown in Fig. 5, we can observe that replies are slightly more effective than quotations and both of these outperform pure content similarity. In other words, the importance of comments can be well evaluated by the logic organization of these comments. Quoting and replying reveal readers' concerns with discussion topics. We also notice that the incorporation of content similarity decreases the system effectiveness. This may seem to contradict our intuition that the textual information should complement the logic-based models. By further investigating our results, we find that content similarity sometimes misleads the decision on the importance of the comments. Besides, the computation cost of calculating the content similarity matrix M_C is very high. Therefore, we only apply the structural information to determine the importance of each comment.

5.8. Recommendation interpretation

We next investigate the precision of our interpretation module. To do that, we focus on a set of 450 relevant articles recommended off 45 topics. These recommended articles have been labeled with specialization, generalization, and duplicate prior to our tests manually. The precision of our interpretation module is summarized in Table 4. We can see that the overall accuracy is 88.7% in our test cases.

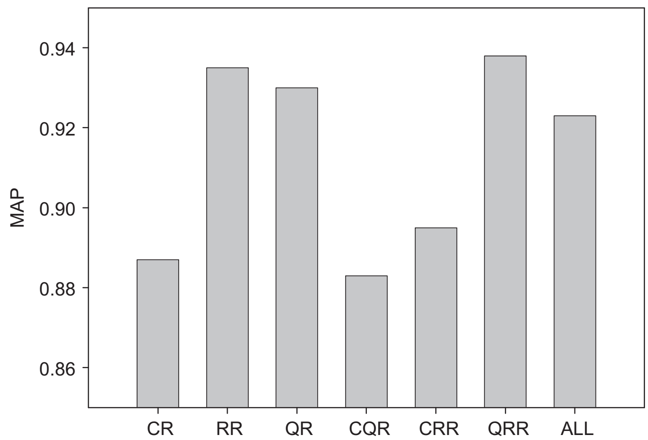

Fig. 5. Effect of semantic and structural relation.

Table 4
Interpretation precision.

	Correct	Tested	Precision (%)
Duplicate	18	18	100
Generalization	189	201	94.0
Specialization	192	231	83.1
Total	399	450	88.7

6. Conclusion and future work

The Web has become a platform for information exchange and user interaction besides information dissemination. Many Web applications are also being extended in this fashion. News recommendation is one such example. Traditional information recommendation is essentially a push service to provide information according to the profile of individual or groups of users. Its niche at the Web 2.0 era lies in its ability to facilitate online discussion by providing relevant news articles to the active participants. In this work, we present a framework for news recommendation in social media that incorporates information from the entire discussion thread. This framework models the logic structure among the comments by distinguishing the cases of replies and quotations. By combining such logic structural information with traditional statistical language models, it can recommend news articles that meet the dynamic nature of a discussion forum. Furthermore, for ease of user consumption of recommended articles, we label each returned article as specialization, generalization, or duplicate. Our tests indicate that this framework is able to return a wider scope of innovative articles from the Web which reflect the trend in the ongoing discussion among its users.

This study can be extended in a few interesting ways. For example, we can use this technique to process personal Web blogs and email archives. The technique itself can also be extended by incorporating such information as reader scores on comments, chronological information of comments, and reputation of users. Indeed, its power is yet to be further improved and investigated.

Acknowledgments

This work has been supported by the Scientific Research Starting Foundation for Returned Overseas Chinese Scholars, the National Natural Science Foundation of China (NSFC) (Grant No. 60803106), the Fok Ying-Tong Education Foundation, China (Grant No. 121068), and the Discovery Grants of Natural Sciences and Engineering Research Council (NSERC) of Canada (No. 303958-2010).

References

[1] J.-W. Ahn, P. Brusilovsky, J. Grady, D. He, S.Y. Syn, Open user profiles for adaptive news systems: help or harm? in: Proceedings of the 16th International Conference on World Wide Web (WWW), 2007, pp. 11–20.
[2] J. Allan, V. Lavrenko, R. Swan, Explorations within Topic Tracking and Detection, Topic Detection and Tracking: Event-based Information Organization, Kluwer Academic Publishers, 2002, pp. 197–224.
[3] R. Baeza-Yates, B. Ribeiro-Neto, Modern information retrieval, Addison Wesley Longman Publisher, 1999.
[4] T. Bogers, A. Bosch, Comparing and evaluating information retrieval algorithms for news recommendation, in: Proceedings of ACM Conference on Recommender Systems, 2007, pp. 141–144.
[5] S. Brin, R. Motwani, L. Page, T. Winograd, What can you do with a Web in your pocket, in: Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 1998, pp. 37–47.
[6] I. Cantador, A. Bellogin, P. Castells, Ontology-based personalized and context-aware recommendations of news items, in: Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI), 2008, pp. 562–565.
[7] J.-H. Chiang, Y.-C. Chen, An intelligent news recommender agent for filtering and categorizing large volumes of text corpus, International Journal of Intelligent Systems 19 (3) (2004) 201–216.
[8] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes, M. Sartin, Combining content-based and collaborative filters in an online newspaper, in: Proceedings of the ACM SIGIR Workshop on Recommender Systems, 1999, pp. 79–91.
[9] A.S. Das, M. Datar, A. Garg, S. Rajaram, Google news personalization: scalable online collaborative filtering, in: Proceedings of the 16th International Conference on World Wide Web (WWW), 2007, pp. 271–280.
[10] G.M. Del Corso, A. Gullí, F. Romani, Ranking a stream of news, in: Proceedings of the 14th International Conference on World Wide Web (WWW), 2005, pp. 97–106.
[11] Y. Guo, Z.Q. Shao, N. Hua, Automatic text categorization based on content analysis with cognitive situation models, Information Sciences 180 (5) (2010) 613–630.
[12] D. Hull, Using statistical testing in the evaluation of retrieval experiments, in: Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1993, pp. 329–338.
[13] G. Karypis, Evaluation of item-based Top-N recommendation algorithms, in: Proceedings of the 10th International Conference on Information and Knowledge Management (CIKM), 2001, pp. 247–254.
[14] H.-J. Lai, T.-P. Liang, Y.C. Ku, Customized internet news services based on customer profiles, in: Proceedings of the Fifth International Conference on Electronic Commerce (ICEC), 2003, pp. 225–229.
[15] V. Lavrenko, W.B. Croft, Relevance based language models, in: Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2001, pp. 120–127.
[16] V. Lavrenko, M. Schmill, D. Lawrie, P. Ogilvie, D. Jensen, J. Allan, Language models for financial news recommendation, in: Proceedings of the Ninth International Conference on Information and Knowledge Management (CIKM), 2000, pp. 389–396.
[17] H.J. Lee, S.J. Park, MONERS: a news recommender for the mobile Web, Expert Systems with Applications 32 (1) (2007) 143–150.
[18] S.K. Lee, Y.H. Cho, S.H. Kim, Collaborative filtering with ordinal scale-based implicit ratings for mobile music recommendations, Information Sciences 180 (11) (2010) 2142–2155.
[19] T. Leek, R. Schwartz, S. Sista, Probabilistic Approaches to Topic Detection and Tracking, Topic Detection and Tracking: Event-based Information Organization, Kluwer Academic Publishers, 2002, pp. 67–83.
[20] D.R. Liu, C.H. Lai, W.J. Lee, A hybrid of sequential rules and collaborative filtering for product recommendation, Information Sciences 179 (20) (2009) 3505–3519.
[21] X. Liu, C. Chen, Y. Liu, Algorithm for ranking news, in: Proceedings of the Third International Conference on Semantics, Knowledge and Grid, 2007, pp. 314–317.
[22] J.H. Liu, P. Dolan, E.R. Pedersen, Personalized news recommendation based on click behavior, in: Proceeding of the 14th International Conference on Intelligent User Interfaces (IUI), 2010, pp. 31–40.
[23] L.N. Osborne, Topic development in USENET newsgroups, Journal of the American Society for Information Science 49 (11) (1998) 1010–1017.

[24] J.M. Ponte, W.B. Croft, A language modeling approach to information retrieval, in: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1998, pp. 275–281.
[25] J. Qiu, L. Liao, P. Li, News recommender system based on topic detection and tracking, in: Proceedings of the Fourth Rough Sets and Knowledge Technology, 2009, pp. 690–697.
[26] S.E. Robertson, S.G. Walker, Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval, in: Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1994, pp. 232–241.
[27] M. Sanderson, J. Zobel, Information retrieval system evaluation: effort, sensitivity, and reliability, in: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2005, pp. 162–169.
[28] J. Su, B. Wang, C. Hsiao, V.S. Tseng, Personalized rough-set-based recommendation by integrating multiple contents and collaborative information, Information Sciences 180 (1) (2010) 113–131.
[29] W. Wang, I. Benbasat, Recommendation agents for electronic commerce: effects of explanation facilities on trusting beliefs, Journal of Management Information Systems 23 (4) (2007) 217–246.
[30] Y. Yang, J.G. Carbonell, R.D. Brown, T. Pierce, B.T. Archibald, X. Liu, Learning approaches for detecting and tracking news events, IEEE Intelligent Systems 14 (4) (1999) 32–43.
[31] X. Yi, J. Allan, A comparative study of utilizing topic models for information retrieval, in: Proceedings of the 31st European Conference on Information Retrieval (ECIR), 2009, pp. 29–41.
[32] J. Zobel, How reliable are the results of large-scale information retrieval experiments? in: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1998, pp. 307–314.
[33] H.T. Zheng, B.Y. Kang, H.G. Kim, Exploiting noun phrases and semantic relationships for text document clustering, Information Sciences 179 (13) (2009) 2249–2262.