Model News Relatedness through User Comments

Xuanhui Wang, Jiang Bian, Yi Chang, Belle Tseng
Yahoo! Labs
Sunnyvale, CA
{xhwang, jbian, yichang, belle}@yahoo-inc.com

ABSTRACT

Most of previous work on news relatedness focuses on news article texts. In this paper, we study the benefit of user-generated comments on modeling news relatedness. Comments contain rich text information which is provided by commenters and rated by readers with thumb-up or thumb-down, but the quality of individual comments varies widely. We compare different ways of capturing relatedness by leveraging both text and user interaction information in comments. Our evaluation based on an editorial data set demonstrates that the text information in comments is very effective to model relatedness while community rating is quite predictive of the comment quality.

Categories and Subject Descriptors: H.3.3 [Information Storage and Retrieval]: Information filtering

General Terms: Algorithms

Keywords: Relatedness, comments, ratings

1. INTRODUCTION

Many online news providers, such as Yahoo! News and New York Times, have allowed users to post comments on published news articles. A popular article can easily attract thousands of comments in a very short period of time. While comments obviously contain rich text and user interaction information, how to unlock the values of comments to benefit various applications has just become an important and promising research direction [4].

In this paper, we explore both text and user interaction information in the application of related news recommendation. Recommending related news articles aims to engage with users after they read current ones and can improve user retention for news providers. Most of existing work such as [3] mainly relies on the news contents to identify related news. Thus, the relatedness captured by these methods are primarily from the viewpoints of *authors*. On the other hand, user comments consist of a set of participants and their expressed textual opinions after reading the articles. Therefore, the viewpoints in comments are from *readers* and can be potentially complementary to those of *authors*. Our goal is to investigate how to effectively use comments to model news relatedness.

Indeed, comments have been actively explored in some recent work (e.g., document summarization [2], Youtube video categorization [1], and cross-media retrieval [4]). However, to the best of our knowledge, there is little work studying how to leverage com-

ments in news relatedness, and thus the benefit of comments for this application is still not well-explored. To address this problem, a few interesting research questions raise: (1) Is the co-commenting behavior well correlated with news relatedness? (2) Given that comments are informally formatted, how useful is the text information? (3) How to combine both community ratings and texts together to reduce the noise?

We aim to answer the above questions in this paper. Based on an editorial data set with human judgments, we design different types of features and examine their power in predicting news relatedness. A few interesting observations are obtained in our study: (1) Due to sparseness of the participants, features based on co-commenting are far inferior than simple text features; (2) The amount of comments varies dramatically over articles and the proper normalization yields a significant impact on the utility of comments; (3) Community rating is predictive of selecting high quality comments to model news relatedness.

2. METHODS

Most of online news services provide the following commenting functionalities: after a user reads an article, the reader can post a small piece of text to comment the article; all the comments (except those abusive ones) are visible to other users and can be rated by them with thumb-up (means like) or thumb-down (means dislike). A few news providers also allow users to recursively reply to existing comments. In this paper, we only consider those comments which are directly attached to the news articles. We will take advantage of the following information to model the relatedness between articles by using *cosine* similarity in the vector space model.

(1) Co-commenting: For each article, we have a set of commenters who post comments to the articles. Our hypothesis is that two articles are related if they share many common commenters. To model this, we have a vector of commenters for each article and use cosine to measure the relatedness (denoted as Co-CM).

(2) Co-rating: Similar to co-commenting, we can define co-thumbing-up (denoted as Co-TU) and co-thumbing-down (denoted as Co-TD) for two articles based on the hypothesis that two articles are similar if their comments are rated by many common raters.

(3) Comment text similarity: The above two approaches use the user interaction information. For comment text similarity, we compute a TF-IDF vector for each comment and use the average vector of all the comments attached to an article as its vector representation. We use the raw term frequency and document frequency is based on the number of comments that a term occurs. The similarity is then computed as the cosine of the two average vectors (denoted as Comment-Text).

There are two problems of the simple text similarity. The first one is that the amount of comments for different articles can vary

Co-CM	Co-TU	Co-TD	Comment-Text
0.1354	-0.0071	0.0133	**0.4181**

Table 1: User interaction vs text features

largely. Some articles have thousands of comments while most of articles only have a few. The second problem is that the quality of comments can vary from spammy or abusive to informative contents. Both of these problems can make the cosine measurement ineffective. Our basic idea is to select a few high quality comments and use this subset to model relatedness instead of using all of them. We explore the following criteria for comment ranking.

- Random: the comments are ranked randomly.
- Length: the longest comments are ranked on the top.
- Oldest: the oldest comments are ranked on the top.
- Newest: the newest comments are ranked on the top.
- Thumb: the comments are ranked based on thumb-up ratio,

$$\frac{T_{up} + a}{T_{up} + T_{down} + b},$$

where T_{up} and T_{down} are the number of thumb-up and thumb-down received from the raters; a and b are smoothing parameters to penalize small sample sizes ($a = 1$ and $b = 10$ in our experiment).

For all of these methods, we select the top n comments and compute the similarity based on average TF-IDF vectors as in the Comment-Text method.

3. EXPERIMENTS

Setup: We use a data set with editorial judgments that has been used in our previous work [3]. We only retain those articles which we can find comments in the same period of Yahoo! comment data. In total, we have 7K relatedness judgments for about 400 seed articles and thus each seed has around 20 labeled articles. Note that when the judgments are collected, comments are not available to the editors. To evaluate each individual feature f, we use the Kendall's rank correlation τ by comparing a rank ordered by f against the ground-truth rank. A higher positive number means better correlation. In addition, we combine all comment-based features with content-based features defined in [3] and learn a relatedness function. We compare this comment-enhanced function with the content-only rank function using NDCG based on 5-fold cross validation.

Results: Table 1 compares the Kendall's rank correlation of Co-CM, Co-TU, Co-TD, and Comment-Text features. To our surprise, Co-CM works poorly compared with Comment-Text. Furthermore, both Co-TU and Co-TD have very low correlation values. We also found that the Co-CM, Co-TU, and Co-TD values are very sparse. For example, 45% of the pairs have 0 values for Co-CM. On the contrary, Comment-Text works pretty well. A possible reason of this observation is that news relatedness emphasizes on topical relevance, while a user is likely to comment on multiple topics, such as both sports and politics articles.

In Figure 1, we demonstrate the performance of different variants of comment ranking methods compared with the Comment-Text baseline. From this figure, we have the following observations: (1) A simple random ranking of comments can achieve much better results than the Comment-Text baseline. The main reason is that different articles can have widely different number of comments. By using a small set of comments, we significantly reduce the difference between articles, which suggests that normalization is im-

Features	NDCG1	NDCG3	NDCG5	NDCG10
Content	0.8054	0.8121	0.8135	0.8546
Content+Comment	0.8231*	0.8155	0.8241**	0.8579

Table 2: Results on learning to rank. * and ** mean the difference is significant at level 0.1 and 0.05.
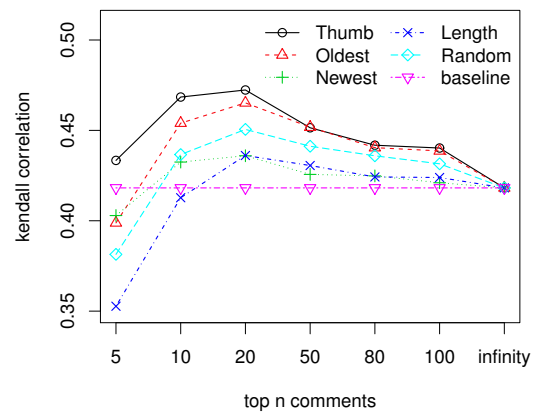

Figure 1: Comparison of different comment ranking methods.

portant when using comments. (2) The Thumb method is the best and can outperform all the other methods over all the range of n values. When $n = 20$, the Thumb method achieves 5% improvement over Random and the t-test shows that the difference is significant (p-value < 0.05). This means that the community rating is very predictive of comment quality. (3) The Oldest variant is the best method among all others except the Thumb one. But the gap between Oldest and Thumb becomes smaller along with the increase of n, which indicates that community rating can differentiate earlier comments better than later ones. Intuitively, earlier comments can attract more ratings and thus the thumb-up ratio can be better estimated.

Finally, we investigate the capability of comment-based features to benefit related news ranking. Table 2 demonstrates the ranking performance in terms of NDCG metrics for ranking function with content features and that with both content and comment based features. From the table, we can find that adding comment-based features can significantly improve the performance of related news ranking, which can further indicate that there are additional complementary information in the comments, and our methods, although preliminary, can effectively explore the benefit of comments.

Discussions: Our preliminary experiments illustrate promising results and there is much potential to improve our methods. For example, more effective ways of combining both interaction and text information can be expected to show better results and thus are interesting for further research.

4. REFERENCES

[1] K. Filippova and K. B. Hall. Improved video categorization from text metadata and user comments. In *SIGIR*, 2011.

[2] M. Hu, A. Sun, and E.-P. Lim. Comments-oriented document summarization: understanding documents with readers' feedback. In *SIGIR*, 2008.

[3] Y. Lv, T. Moon, P. Kolari, Z. Zheng, X. Wang, and Y. Chang. Learning to model relatedness for news recommendation. In *WWW*, 2011.

[4] M. Potthast, B. Stein, F. Loose, and S. Becker. Information retrieval in the commentsphere. *ACM Tran. on Int. Sys. and Tech.*, 2012.