Tweet-Recommender:
Finding Relevant Tweets for News Articles

Ralf Krestel, Thomas Werkmeister,
Timur Pratama Wiradarma
Hasso Plattner Institute
Potsdam, Germany

Gjergji Kasneci
Schufa Holding AG
Wiesbaden, Germany

ABSTRACT

Twitter has become a prime source for disseminating news and opinions. However, the length of tweets prohibits detailed descriptions; instead, tweets sometimes contain URLs that link to detailed news articles. In this paper, we devise generic techniques for recommending tweets for any given news article. To evaluate and compare the different techniques, we collected tens of thousands of tweets and news articles and conducted a user study on the relevance of recommendations.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

Keywords

Recommender Systems, News Articles, Twitter

1. INTRODUCTION

More than half a billion tweets are written every day. Many of these deal with personal issues or status updates of the kind "I am at . . . ", or "I feel like . . . ". But there are also many tweets that comment on current events, express opinions about politics, entertainment, sports, or report new information about an event. The structure of this communication and the discussed topics are similar to traditional journalism [2]. While Twitter is a great source for disseminating short news snippets along with user opinions and moods, traditional newspapers are good at providing facts or detailed information on events.

In this paper we present an approach for combining news articles and tweets to not only cover an event by the facts and detailed information available in news articles, but also by the opinions expressed in related tweets. Imagine a controversial news article being published. Wouldn't it be nice to instantly see the reaction of the social communities on the Web by finding relevant tweets? Or think about politicians who could track opinions on articles of political and social interest, or about journalists who could use such tools to track the sentiments generated by their articles on the Web. By exploring opinions and sentiments of social communities on particular topics, journalists are also able to re-adjust their focus in follow-up articles.

With the above scenarios in mind, we derive various techniques for recommending tweets to news articles. In summary, our contributions are as follows: (1) We devise/train various models for recommending tweets to news articles. (2) In order to adequately train the models, we collected thousands of relevance judgements on a dataset of tens of thousands of tweets and news articles. (3) We evaluated the devised models based on an extensive user study.

2. RELATED WORK

A couple of works focus on comparing twitter with traditional news media. Zhao et al. [3], for example, use topic models to analyze news articles and tweets and show how their topic distributions differ. Sankaranarayanan et al. [2] present a framework to cluster real-time tweets and tag them with geolocations in order to gather breaking news faster than conventional news media. Zubiaga et al. [4] present an application to assist journalists with gathering news based on Twitter. They filter newsworthy tweets using machine learning algorithms based on the content of a tweet along with additional information (hashtags, profile images, etc.). Those works show that on the one hand microblogs contain valuable information comparable to conventional news media, but on the other hand there are also characteristic challenges, such as noise and misleading information that makes finding relevant information a non-trivial task.

The most similar work to ours is presented by Cao et al. [1] who generate comments for Chinese news by mining microblog posts. In contrast to their work, we focus on diverse tweets that add information to a given news article.

3. RECOMMENDING TWEETS

Finding relevant tweets for news has various challenges: (a) the tweets need to be of similar content as the news article, (b) the tweets need to be published at around the same time as the article, and (c) the tweets in the recommendation list shouldn't be redundant. To identify similar content tweets we employed language models (to find word overlap) and topic models (to find concept overlap). To combine the resulting content similarity scores with other features, such as recency or popularity of a tweet, we used logistic regression as well as boosting.

Language Models. Similar to traditional information retrieval tasks we can use language models to rank tweets based on news articles. The news article can be seen as the query and the tweets are the document collection. Instead of a query likelihood model which assumes that the queries are short and the documents are long, we use a document likelihood model to compute the probability that a tweet is generated from a news article. We smooth our language model using Dirichlet smoothing. We addressed tweet length normalization by calculating the geometric mean of the factors involved. A simple threshold on the document likelihood was then used to predict relevance.

Topic Models. We employ latent Dirichlet allocation and compare the tweets' topics with the news article's topics to find the most relevant tweets. We use a threshold on the score to classify tweets into relevant and non-relevant.

Logistic Regression. To see whether other tweet features can be exploited to find relevant tweets we learnt a logistic regression model. Besides language and topic model scores, we included 16 additional features, such as publication time, length, follower count, etc. Therefore we manually annotated a set of tweets as relevant or non-relevant with respect to 17 news articles.

Boosting. Instead of logistic regression that separates instances linearly, boosting performs a non-linear classification of data. We use AdaBoost as a boosting algorithm and for the weak learners we use decision stumps, a simple type of decision tree that separates data linearly. Boosting combines multiple decision stumps and therefore allows for example the identification of tweets that are similar to a given news article but not too similar.

Diversity. To guarantee diversity within our set of recommended tweets and be confident that the tweets add valuable information to the news article, we need to make sure that (a) a tweet is at least not identical with the headline of the news article; (b) there are no (near) duplicates in the final ranking of tweets. We achieve this by ignoring tweets that have a high word overlap with tweets that are already in the set of recommended tweets. Because we couldn't extract the headlines reliably from the news article, we removed tweets consisting only of the headline and a link to the news article by hand in our test set. More sophisticated ways for automatic detection of near duplicates could be thought of, as well as of automatic ways to identify the headline of a news article and making sure that a tweet contains more than just that.

4. EVALUATION

The goal of our tweet recommender is to find relevant tweets for news. To evaluate its performance and to train the machine learning models we require relevance judgements for a set of tweets with respect to a set of news articles. Due to the lack of existing test collections we conducted a user study get a set of training and test tweets with relevance judgments.

Dataset. Our dataset was collected from Twitter between July and November 2013. We only kept English tweet containing at least one URL and at least one predefined hashtag. The hashtags were chosen based on a seed set of hashtags consisting of #obama, #snowden and #merkel. It was then extended with co-occurring hashtags such as #president, #pjnet, #jobs, #teaparty, #politics, #auspol, #surveillance, #germany, #syria etc. After filtering for ac-

Table 1: Performance of Different Methods

Method	Accuracy
Language Model	64.1%
Topic Model	77.8%
Logistic Regression	80.8%
Boosting	82.5%

tual news articles 55k Web pages and 121k tweets formed the final dataset.

User Study. To evaluate our system we conducted a user study in which each participant had to judge whether a tweet is relevant with respect to a given news articles. A tweet should be marked as relevant if it (a) adds information, (b) fits to the broader topic of the news article, (c) confirms the information of the news article, (d) contains an interesting opinion, or (e) backs up the presented opinion of the article by an argument.

The study was conducted using a Web interface. Each user was presented a news article and a selection of 20 tweets. The task was to read the news article and to classify the tweets as relevant or not. For this study we handpicked 17 news articles covering different topics. For each article 20 tweets were shown to the user. Those 20 tweets included 10 tweets from the language model ranking and 10 tweets from the topic model ranking. In the user study we collected 1602 relevance judgements from 11 computer science students.

Results. We compared the accuracy of our four different recommender systems. We picked the associated tweets of 4 randomly selected news articles for testing and used the tweets of the remaining 13 news articles for training. Table 1 shows the result on the test set.

Regarding logistic regression, the highest weight was learnt for the topic model score followed by publication time of the tweet. Other features such as length of a tweet, followers count of the tweet's author, or number of hashtags in a tweet had no significant influence in the model. Note that tweets containing the URL of the actual news article are not necessarily relevant, e.g. if they only repeat the headline of the article and don't add any more value.

5. CONCLUSIONS

We compared different approaches to recommend tweets for news articles. Using topic models to find similar tweets outperformed the language model approach and we showed that adding more features and combining them using logistic regression or AdaBoost further improves the results.

6. REFERENCES

[1] X. Cao, K. Chen, R. Long, G. Zheng, and Y. Yu. News comments generation via mining microblogs. In *WWW*, pages 471–472. ACM, 2012.

[2] J. Sankaranarayanan, H. Samet, B. E. Teitler, M. D. Lieberman, and J. Sperling. Twitterstand: News in tweets. In *GIS*, pages 42–51. ACM, 2009.

[3] W. X. Zhao, J. Jiang, J. Weng, J. He, E.-P. Lim, H. Yan, and X. Li. Comparing twitter and traditional media using topic models. In *ECIR*, pages 338–349. Springer, 2011.

[4] A. Zubiaga, H. Ji, and K. Knight. Curating and contextualizing twitter stories to assist with social newsgathering. In *IUI*, pages 213–224. ACM, 2013.