Recommendation for Online Social Feeds by Exploiting User Response Behavior

Ping-Han Soh
Graduate Inst. of
Communication Engineering
National Taiwan University
Taipei, Taiwan
phsoh@arbor.ee.ntu.edu.tw

Yu-Chieh Lin
Dept. of Electrical Engineering
National Taiwan University
Taipei, Taiwan
yuccalin@arbor.ee.ntu.edu.tw

Ming-Syan Chen
Research Ctr. for Information
Technology Innovation
Academia Sinica
Taipei, Taiwan
mschen@citi.sinica.edu.tw

ABSTRACT

In recent years, online social networks have been dramatically expanded. Active users spend hours communicating with each other via these networks such that an enormous amount of data is created every second. The tremendous amount of newly created information costs users much time to discover interesting messages from their online social feeds. The problem is even exacerbated if users access these networks via mobile devices. To assist users in discovering interesting messages efficiently, in this paper, we propose a new approach to recommend interesting messages for each user by exploiting the user's response behavior. We extract data from the most popular social network, and the experimental results show that the proposed approach is effective and efficient.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces

Keywords

social networks; social feeds; recommender systems

1. INTRODUCTION

There are a huge number of messages posted on online social networks such as Facebook and Twitter. On Facebook, it would take a user quite much time to read all the messages posted by the user's friends. To save time, a user would like to first read interesting messages and then read or skip the rest. Thus, it is appreciated if interesting messages are displayed first. It is worth noting that recommendation focusing on Twitter data [7] cannot be directly applied to Facebook data due to different scenarios.

A naive approach is to filter the messages based on a user-defined friend list. This is not practical since it requires each user to know all friends well and be conscious of friends' change. Another approach is to use recommendation systems available. However, content-based recommendation systems are not appropriate since most messages posted on online social networks are short and even ambiguous. Besides, collaborative filtering [2] or graph-based recommendation [5] is not practical since the number of messages is very large but users respond to only quite a small ratio of the messages. It is worth noting that some recent content-based recommendation systems consider the trust [3] [4] or social information [6] to improve the recommendation results. This suggests that considering the relationship between users can improve the recommendation results.

In this work, we propose to recommend interesting messages by exploiting user response behavior, where a user responds to a message is regarded as the message is interesting to the user. Consider the following two types of response behaviors. A user responds to a message because she likes the message itself; a user responds to a message because she sees several friends have already responded to it. For the first, we extract a categorical feature set about the message itself such as creator, destination, and type. For the second, we extract the user set that have already responded to the message. With the two kinds of information, we design a score to jointly consider the two types of response behaviors. Messages ranked by the score are ready for display. It is worth noting that Freyne's work [1] can generate several ranking lists of feeds based on different relevance judgements. However, our goal is to provide a single ranking list by jointly considering the two response behaviors.

WWW 2013 Companion, May 13–17, 2013, Rio de Janeiro, Brazil.
ACM 978-1-4503-2038-2/13/05.

2. METHODOLOGY

We define the social feed score F as a linear combination of two conditional probabilities, which are corresponding to the two response behaviors. Given an unread message of a user, the *social feed score* is defined as

$$F = (1 - \mu) \times P_M + \mu \times P_R,$$

where P_M is corresponding to the *message feature set* and P_R is corresponding to the *responded user set*. In other words, P_M is the probability that the user responds to the message conditional on the given message feature set while P_R is the probability conditional on the given responded user set. Let R denote the responded user set and let Y denote the user's response, we define

$$P_R = 1 - \prod_{\forall r \in R} (1 - P(Y|r)),$$

where $P(Y|r)$ can be learnt from the response history. The user has responded to some of the messages that the user can see and are with user r responded. The ratio of the responded ones to all is $P(Y|r)$. The P_M is defined in the same way.

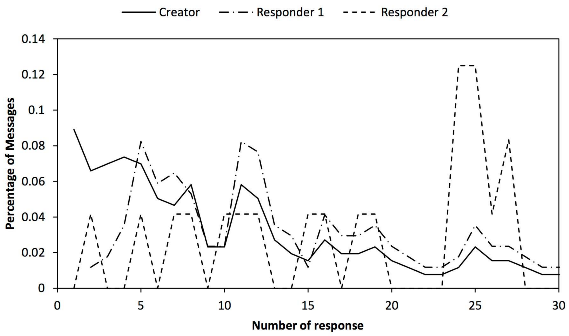

Figure 1: Distributions of the number of responses.

It is important to find a good μ value for balancing P_M and P_R. The μ value is decided by observing which response behavior is more likely to happen. We explain the idea with the example in Fig. 1. For a message creator, the number of responses of each message she creates is summarized as a distribution, which is the line labeled as *Creator*. Among the messages created by the same creator, we find the messages that have been responded by a user and summarize the number of responses of the messages as a distribution, which is the dotted line labeled as *Responder* 1 or *Responder* 2. It is observed that the distribution of Creator is basically decreasing. However, Responder 2's distribution is quite different from Creator's. Moreover, Responder 2's distribution implies that Responder 2 is inclined to respond to the messages that are already responded by many other users. On the contrary, Responder 1's distribution is similar to Creator's, which implies that the response behavior is mainly based on the message feature set but not on the responded user set. Based on the above, we propose to decide the μ value by comparing the response distribution of the user and the message creators using the Hellinger distance, which is 0 if two distributions are identical. To avoid overfitting, we regard μ as the average Hellinger distance between the user's distribution and each creator's distribution.

With P_M, P_R, and μ, the F score of a message can be obtained. Messages are then ranked by F and the results are the recommended social feeds. It is worth noting that the effect of time decay can be applied to F easily by multiplying F by a time decaying coefficient.

3. EXPERIMENTS

The experimental dataset is extracted from Facebook using Facebook Graph API. We collect all social feeds of 44 active users during the period from 2012/02/01 to 2012/06/17. The *response* is referred to that a user *likes* a message. Social feeds during 2012/02/01 and 2012/06/08 are used for calculating P_M, P_R, and μ, and the rest are regarded as unread messages. Without losing generality, we use the message creator as the message feature for calculating P_M.

We first examine the effectiveness of the proposed method. For each user, we rank the unread messages based on the proposed F score, which uses the average Hellinger distance as μ, and we display the top-20 messages as the *recommendation list*. Besides, we also generate a recommendation list based on only the message feature set ($\mu = 0$) and a recommendation list based on only the responded user set ($\mu = 1$).

Table 1: Experimental results.

Method	Precision
Facebook Newsfeed	0.0330
Message feature set only ($\mu = 0$)	0.1217
Responded user set only ($\mu = 1$)	0.1366
Proposed F (avg. Hellinger dist.)	0.1426
Ideal result	0.2642

Table 1 shows the results. We measure the recommendation lists by precision, which is the number of user-responded messages over 20. Precision of the Facebook Newsfeed is also provided for reference. The proposed method outperforms the others by 17.2% and 4.4%. This shows that it is appropriate to jointly consider the message feature set and the responded user set. It is worth noting that the ideal precision is only 0.2642, which means that a user averagely responds to 5.284 messages among the unread messages.

We perform the calculation of P_M, P_R, and μ in advance. To provide the recommendation list online, we efficiently calculate the feed score F of each unread message using the pre-calculated P_M, P_R, and μ values, and then we rank the unread messages based on the score. In our experiments, online calculations cost less than 60ms.

4. CONCLUSION

In this paper, we propose a new approach to recommend interesting social feeds by exploiting user response behavior. We define the social feed score F to jointly consider two kinds of response behaviors, corresponding to the message feature set and to the responded user set. To balance the two effects, we decide the μ value by comparing the response distributions of the user and creators. The social feeds are generated by ranking the unread messages based on F score. The experiments are performed on a real dataset, and the experimental results show that the proposed approach is effective and efficient.

5. REFERENCES

[1] J. Freyne, S. Berkovsky, E.M. Daly, and W. Geyer. Social networking feeds: recommending items of interest. In *Proc. of RecSys*, 2010.

[2] Y. Koren. Collaborative filtering with temporal dynamics. In *Proc. of KDD*, 2009.

[3] H. Ma, I. King, and M.R. Lyu. Learning to recommend with social trust ensemble. In *Proc. of SIGIR*, 2009.

[4] H. Ma, D. Zhou, C. Liu, M.R. Lyu, and I. King. Recommender systems with social regularization. In *Proc. of WSDM*, 2011.

[5] L. Xiang, Q. Yuan, S. Zhao, L. Chen, X. Zhang, Q. Yang, and J. Sun. Temporal recommendation on graphs via long- and short-term preference fusion. In *Proc. of KDD*, 2010.

[6] X. Xin, I. King, H. Deng, and M.R. Lyu. A social recommendation framework based on multi-scale continuous conditional random fields. In *Proc. of CIKM*, 2009.

[7] L. Yang, T. Sun, M. Zhang, and Q. Mei. We know what @you #tag: does the dual role affect hashtag adoption? In *Proc. of WWW*, 2012.