Unseen Tweet Recommendation via Content and Network Analysis ⋆

Noor Aldeen Alawad[1], Aris Anagnostopoulos[1], Stefano Leonardi[1], Ida Mele[2], and Fabrizio Silvestri[3]

(1) Sapienza University of Rome, Italy,
{alawad,aris,leonardi}@dis.uniroma1.it,
(2) University of Lugano, Switzerland,
ida.mele@usi.ch
(3) Yahoo! Research, London, UK,
silvestr@yahoo-inc.com

Abstract. We present a novel approach for recommending concealed tweets, originally proposed in [1]. The tweets we recommend are potentially interesting for the user, but cannot be seen by the user, because nobody in the her social circles published or retweeted them. Our methodology leverages the structure of the network with two-hop distance from the user, the content similarity, and the analysis of shared retweets in order to discover the interesting tweets to recommend, which otherwise would remain hidden from the user.

Keywords: Tweet Recommendation, Content and Network Analysis.

1 Introduction

Twitter is a popular micro-blogging system where users can post short messages, called tweets, consisting of personal information, status updates, news, or links to webpages and other web content. Given the rapid proliferation of Twitter, its users often feel overwhelmed with information, and tweet recommendation is an important service which helps these users to discover potentially interesting tweets.

Most of Twitter's recommendation systems are based on recommending users to follow [3], webpages to visit [7], or they consist in a reranking of the tweets that appear in the user's timeline [6]. Recent research is based on recommending concealed tweets (i.e., tweets that are not posted or retweeted by anybody in the user's social circles). Consider, for example, a user who is interested in computer science and everyday she checks the tweets of the computer scientists she is following. Some of these tweets are originally posted by them, whereas others are just the result of retweetting actions. When nobody has retweeted a message that is potentially interesting for the user, such message remains concealed from the user's eyes, because it does not appear in her timeline. Pennacchiotti et al. [4] tackled this problem by proposing tweets whose content matches the user's interests. In [1] we extended their idea of recommending hidden tweets that are potentially interesting to the user, but instead of using only the content

⋆ This paper is an extended abstract of a paper that has been accepted as a short paper to SIGIR 2016.

analysis, we proposed to exploit the structure of the network around the user as well as the analysis of mutual retweets, and we demonstrated that this yields an improvement in the quality of the recommendations.

2 Methodology

We aim at recommending tweets hidden from the user, because, for example, nobody in her circles wrote or retweeted them. As a result, Twitter users who subscribe to our system can receive recommendations of novel tweets, which are of potential interest, but they do not appear in their timeline and would be missed.

For making recommendations we applied network, content, and retweet analyses. The idea is to use the network structure around the user to find candidate tweets to recommend. Then, these tweets are ranked by using the content-similarity features and the number of common retweets, which is an indication of how much the users' interests are similar.

More in details, when a user u subscribes to the service, the system uses the Twitter API to retrieve u's friends, friends of friends, and timeline (i.e., tweets and retweets). Given u's ego-network up to depth two, the recommendation algorithm exploits the transitivity property of the following–follower relationships. For example, assume that user u follows a set of users v_i who follow user z, and that u does not follow z, the (re)tweets of z are not visible to u, unless some v_i retweeted them. The idea is to take z's (re)tweets that do not appear in u's timeline and use them as possible candidates for the recommendations. To weigh the importance of the tweets of z, we count the number of users v_i that are in between.

Network Analysis. The scenario described above can be mapped to the problem of finding open triangles in the ego-network of a user. Inspired by the MapReduce approach of Suri et al. [5] for counting triangles, we designed and implemented a MapReduce algorithm to find open triangles, so that when user u follows v, and v follows z, the predicted link would be (u, z), and recommendations would go from z to u as if u is actually one of z's followers. We also count the missing edges that close triangles in order to rank the nodes at distance two from the ego based on how many incoming links they have. Specifically, given the ego node u, let $\Gamma(u)$ be the set of u's friends (followees) and $\Gamma(\Gamma(u)) \setminus \Gamma(u)$ the set of u's friends of friends who are not friends of u. We define as the weight of user $z \in \Gamma(\Gamma(u)) \setminus \Gamma(u)$ to be the number of in-links $weight(z) = |(\Gamma(u), z)|$. Nodes are then ranked based on decreasing values of their weights.

Content Analysis. Tweets that best match the user's interests can be discovered by applying the content similarity between a candidate tweet and the ego's (re)tweets as well as the similarity between ego's and candidate users' timelines.

As content-similarity measures we used the *cosine similarity* and *jaccard distance*. We computed the similarity of tweets by using single terms and bi-grams. One type of similarity is computed between the timelines of u and of $z \in \Gamma(\Gamma(u))$ and another similarity is between the candidate tweet originated from z and u's (re)tweets. This gives us an overall number of six content-similarity features: both cosine similarity and jaccard distance for the tweets, the timelines, and the timelines using bigrams.

Retweet Analysis. Our recommendation algorithm includes another feature based on the number of common retweets, which provides an indication of the similarity between the interests of the user u and her neighbors at distance two. Indeed, we noticed that users share on average 15 retweets, so we can use the number of mutual retweets to infer how close are the users' interests [2].

Ranking of Recommendations. We obtained the ranked list of tweets to recommend by using the pairwise comparison. In more detail, we considered all pairwise combinations of candidate tweets and we compared them with respect to the features. We assumed that a tweet beats another one if it has a better value for more features. At the end each tweet has a number of wins against the rest of the tweets, and the tweets with higher number of wins are more likely to be relevant to the user's interests, hence they are shown on the top of the ranked list of recommendations.

3 Experimental Results

To validate our methodology, we conducted a user study where we proposed to real Twitter users our recommended tweets and collected their feedback. We could not use retweets to assess if the recommendations were interesting or not, because we our approach aims at recommending tweets that are not visible to the users, hence they cannot have been retweeted by them.

User Study Evaluation. In the user study were involved 42 active Twitter users. The users could participate to our experiment by registering to our system, which retrieves the user's information needed to make the recommendations, and, once these are ready, it notifies the user, who can see and rate the recommended tweets. As a competitor we used the approach presented by Pennacchiotti et al. [4], which applies the content similarity among tweets. As our baseline we adopted their approach based on the cosine similarity. In particular, we proposed the top-5 recommended tweets from our approach and the top-5 recommendations from the baseline. The two rankings were presented to the users, in a way that they could not identify what system was used for creating the corresponding ranking. Following the same experiment of [4], our users could rate the proposed recommendations using a four-grade scale: *Excellent:* the tweet is very interesting/informative w.r.t. her interests, *Good:* the tweet is interesting/informative w.r.t. her interests, *Fair:* the tweet is somehow interesting/informative w.r.t. her interests, and *Bad:* the tweet is not interesting/informative at all.

Assessing the Performance of the Recommendation System. For assessing the performance of our recommendation algorithm, we computed the *precision*$@k$ (p@k), which is the percentage of relevant tweets found in the top-k ranked tweets, and *normalized Discounted Cumulative Gain* (nDCG), which measures the performance of a recommendation system based on the graded relevance of the recommendations. We calculated the average of these metrics over all the users who subscribed to the system, and, compared to the baseline, we could observe an average improvement of 12.4% for the precision and of 1.6% for the nDCG metric. We computed also the *Reciprocal Rank* (RR), which is the inverse of the ranking position of the first relevant tweet and the average over all users is of 91% for our approach and 83% for the baseline. In Figure 1 is shown the Likert scale for both methods. As we can observe there is a larger number of tweets rated as *Excellent* and *Good*, while the number of tweets rated as *Bad* is lower compared to the baseline.

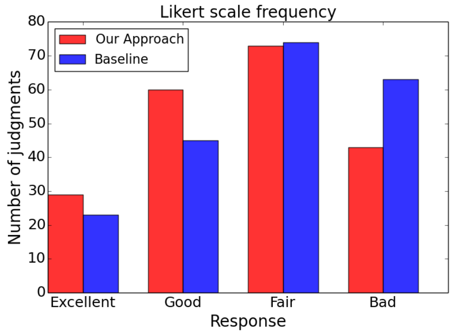

Fig. 1. Comparison between our approach (red) and the baseline (blue) for the Likert scale.

4 Conclusions

We have presented an approach for recommending concealed tweets. We have also summarized the results of the user study, involving real Twitter users, which we conducted to evaluate our methodology. The results corroborated that our system is able to overcome the existing approach for unseen-tweet recommendation. In particular, we could observe an improvement of 12.4% in the precision compared to the baseline.

References

1. N. A. Alawad, A. Anagnostopoulos, S. Leonardi, I. Mele, F. Silvestri: Network-Aware Recommendations of Novel Tweets. In the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval. *To appear*. 2016
2. K. Chen, T. Chen, G. Zheng, O. Jin, E. Yao, Y. Yu: Collaborative personalized tweet recommendation. In the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 661-670. ACM (2012)
3. J. Hannon, K. McCarthy, B. Smyth.: Finding useful users on Twitter: Twittomender the followee recommender. In the 33rd European Conference on Information Retrieval, pp. 784–787. Springer Berlin Heidelberg (2011)
4. M. Pennacchiotti, F. Silvestri, H. Vahabi, R. Venturini: Making your interests follow you on twitter. In the 21st ACM International Conference on Information and Knowledge Management, pp. 165-174. ACM (2012)
5. S. Suri and S. Vassilvitskii: Counting triangles and the curse of the last reducer. In the 20th International Conference on World Wide Web, pp. 607-614. ACM (2011)
6. R. Yan, M. Lapata, X. Li: Tweet recommendation with graph co-ranking. In 50th Annual Meeting of the Association for Computational Linguistics, pp. 516-525, Association for Computational Linguistics (2012)
7. N. Yazdanfar and A. Thomo: Link recommender: Collaborative-Filtering for recommending URLS to Twitter users. Procedia Computer Science. 19, pp.412-419 (2013)