Finding Useful Users on Twitter: Twittomender the Followee Recommender*

John Hannon, Kevin McCarthy, and Barry Smyth

CLARITY: Centre for Sensor Web Technologies,
School of Computer Science & Informatics,
University College Dublin, Ireland
{firstname.surname}@ucd.ie

Abstract. This paper examines an application for finding pertinent friends (followees) on Twitter[1]. Whilst Twitter provides a great basis for receiving information, we believe a potential downfall lies in the lack of an effective way in which users of Twitter can find other Twitter users to follow. We apply several recommendation techniques to build a followee recommender for Twitter. We evaluate a variety of different recommendation strategies, using real-user data, to demonstrate the potential for this recommender system to correctly identify and promote interesting users who are worth following.

1 Introduction

Twitter has proven to be one of the most surprising success stories of recent web history. To get the most from Twitter, users need to follow others so that they can benefit from the tweets of these followees. But who should a user follow, beside their immediate friends? Helping users to find new people to follow is an important challenge and the focus of this paper. Twitters' own recommendation system utilises two main approaches when attempting to create new connections between users. Firstly, they use the categorisation approach, this aims to box-off celebrities or popular users into categories such as Entertainment, Music, etc. for users to select from. This approach, whilst appropriate for initial connections for new users, may become less useful as users engage with the system and find that forming connections is usually easier by reading content from Re-Tweeted (forwarded) tweets from users they are not connected to or from other online resources relevant to them that have associated Twitter accounts. Twitter's second approach uses collaborative filtering techniques. Much research has been carried out using collaborative filtering to aid in filling in the missing links, be it a rating on a movie review site [4] or in this case filling in the links in ones social graph (see also [1, 2]). The main idea behind this approach is that your friend's friend has the potential to be your friend. These approaches alone do not fully allow

* This work is supported by Science Foundation Ireland under grant 07/CE/I1147 and by Amdocs Inc.

[1] http://www.twitter.com

P. Clough et al. (Eds.): ECIR 2011, LNCS 6611, pp. 784–787, 2011.

for the best or most relevant recommendations for a user to be produced and neither does sifting through the Twitter public timeline of thousands of tweets per second in the hope of finding someone of interest. The collaborative approach is good at filling in the missing links, but what happens if we want to find users with similar interests outside of our extended social circle? This is where *Twittomender* [3] comes in, helping its users in finding useful followees, who produce relevant content, and preventing information overload. Here, we examine the proven collaborative style approaches and also look at the content contributed to Twitter – we utilise users tweets as a source of information to compare against other users and produce a content-based recommendation system [5].

2 Twittomender: Recommending Users to follow

Twittomender has been developed as web service[2]. Twittomender mines a large number of user profile details and content (tweets, followee/follower user ids) by following the links between users on the social graph with the aid of Twitter's API[3]. When a new user is looking for followees, Twittomender can recommend profiles from our database of over 1 million Twitter users. Twittomender allows users to input a search query to find interesting potential followees or indeed a search query can be derived implicitly from that users own tweets. The terms of this query relate to the types of content the user would like to consume. These terms are used to find users who have tweeted these key terms frequently. These Twitter profiles are stored as documents and indexed by Lucene[4]. By using Lucene, a document search engine, we can search these documents in an information retrieval style by using queries to find matching documents. Twittomender uses 7 recommendation strategies, these include the content (tweets) and connections (ids) of a user. These fields are queried against the document corpus for similarity. Twittomender's recommendation strategies are comprised of two recommendation categorisations; **Content-based**: *(1) Users own tweets, (2) Followee's tweets, (3) Follower's tweets, (4) All tweets*, and **Collaborative-based**: *(5) Followee's Ids, (6) Follower's Ids, (7) All Ids.*

2.1 System Interaction

A typical user interaction with the system can be seen in Figure 1. Users of Twittomender are first asked to sync their Twitter account with Twittomender so that we can retrieve their profile information and form a user document. Each user in the Twittomender database is modeled as a document, with the documents content containing up to 200 of the most recent tweets of that user. Using the TF-IDF scoring metric, which examines the frequency of a given term within a document multiplied by its frequency score across all the documents in that index, search queries entered are scored for similarity amongst the document

[2] http://twittomender.ucd.ie

[3] http://apiwiki.twitter.com

[4] http://lucene.apache.org/java/docs/

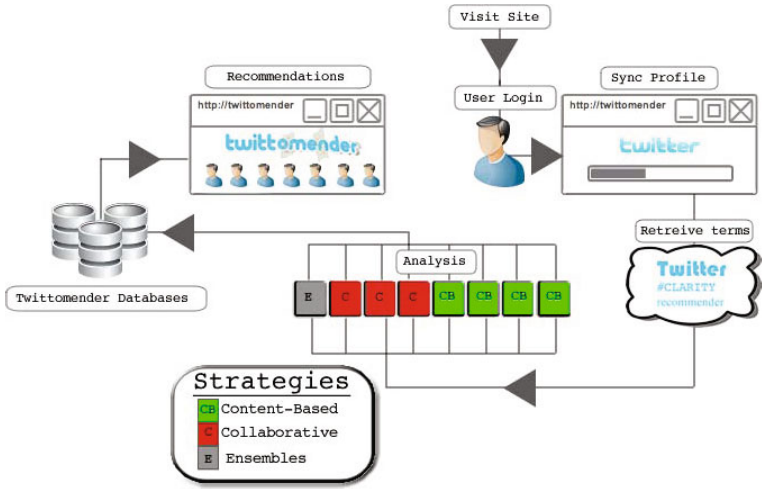

Fig. 1. The Twittomender System Control Flow Diagram

corpus to produce a result set. These search queries are either top terms from that user's tweets (content-based) or their followees/followers ids (collaborative-based). For each strategy the top 20 of these document results are returned to the system as recommendations for followees. To form the final recommendations that are presented to a given user we then combine all our 7 recommendation strategies to form an ensemble approach that represents the most frequent suggestions across all strategies. The users recommended in the ensemble approach are then presented to the synced user with their profile pictures, user statistics and a term cloud of frequent terms from tweets to explain the recommendation and aid in selecting potential followees.

3 Evaluation

To evaluate Twittomender, we carried out two rounds of testing. The initial offline evaluation comprised of selecting 20,000 user documents from our database. This group of documents was split into two sets, a 1000 user test set and a 19,000 user recommendation pool to suggest users from, for each of the test users. Our precision metric for these offline evaluations was centered around whether or not a suggested user was already being followed by that test user, if so, it was then classed as a valid recommendation. There is a downside to this offline approach. How do we discern the validity of the other suggested users that are not being followed by the current test user. The solution to this is an online user study.

In the online study, we asked active users of Twitter to sync their Twitter profiles with Twittomender and each one was given a set of recommendations. The main aim behind this online trial was to find out whether the unknowns from a recommendation set were valid also. To this end we removed recommended users that the synced user was already following to find out the number of unknowns they would follow from the recommendations.

Results from both the offline and online evaluations have proven successful. In the offline evaluations, each test user was presented with a recommendation list of 20 users. On average, across 1000 test users, 5 followees were identified for each user. This success rate is from our best performing strategy, user documents created from a users follower connections (collaborative-based). In the online trial the 34 participants indicated they would be willing to follow on average 6.9 new users from a recommendation list size of 30. These are users that they are not already following. Both these statistics bode well for the potential for Twittomender to recommend followees both old and new.

4 Conclusion

In this paper we have shown the Twittomender system as an effective way of finding new people to follow on Twitter. We have introduced 8 various recommendation strategies employed in Twittomender, namely 4 content, 3 collaborative and the 1 ensemble approach. Clearly the 20,000 users selected for our experiment only represent a snapshot of Twitters users, but we believe they form a diverse section of the Twitter community and future work will expand on these test sets. From recent user trials of the system there is a clear indication as to the efficiency of Twittomender to produce new connections for a user. As future work we aim to extend the live trial to more users and to compare against the recommendations of Twitter's own friend recommender system. Also we plan to examine the nature in which people follow others, as these strategies currently can not predict users who are followed due to some event or activity taking place in the real world.

References

1. Chen, J., Geyer, W., Dugan, C., Muller, M., Guy, I.: Make new friends, but keep the old: recommending people on social networking sites. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI 2009, pp. 201–210. ACM, New York (2009)
2. Guy, I., Ronen, I., Wilcox, E.: Do you know?: recommending people to invite into your social network. In: Proceedings of the 13th International Conference on Intelligent User Interfaces, IUI 2009, pp. 77–86. ACM, New York (2009)
3. Hannon, J., Bennett, M., Smyth, B.: Recommending twitter users to follow using content and collaborative filtering approaches. In: Proceedings of the Fourth ACM Conference on Recommender Systems, RecSys 2010, pp. 199–206. ACM, New York (2010)
4. Konstan, J.A., Miller, B.N., Maltz, D., Herlocker, J.L., Gordon, L.R., Riedl, J.: Grouplens: applying collaborative filtering to usenet news. Commun. ACM 40, 77–87 (1997)
5. Pazzani, M.J., Billsus, D.: Content-based recommendation systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 325–341. Springer, Heidelberg (2007)