Abstract: This paper presents an approach to personalised recommendation of who to follow on Twitter. It is based on a simple observation that following the same Twitter account could be considered a fact that contributes to a possible similarity of followers. We proposed a simple scheme to elicit gradually candidates to follow which are most likely sources of microblogs that are of some interest to the user. We implemented a prototype and conducted experiments. They show that the basic idea works. We compared our system with the known “Who To Follow” application. Our system shows at least comparable performance and often gives better recommendations.

Keywords: Personalised Recommendation, Microblog, Twitter, Information Stream.

1 INTRODUCTION

Information overload is one of the significant characteristics of the present era. It manifests itself in many forms, one important of them being an information stream. The online social networking service Twitter administers the creation and sharing of microblogs called tweets. Users write and post very short (up to 140 characters) messages. Twitter distributes it to other users who subscribed to receiving messages from them by declaring themselves to be their followers. A user can also forward a received tweet (retweeting), which gives them the power to spread information broadly. In such a way, users receive streams of tweets. To put this into context, there are (as of July 2015) 645 million registered Twitter users, out of which 289 million are active ones, generating on average of 58 million tweets per day (Statistic Brain Institute, 2015). Each user faces a question: Which tweets from this huge stream ordered on a timeline should be a part of the specific stream that is shown to them?

A user can influence contents of own stream by subscribing to certain accounts as their follower. If they are too cautious, subscribing to only very few accounts, the user may miss many interesting microblogs. If they follow everyone who potentially writes something interesting, they may be flooded by so many tweets that it becomes impossible to read them all in a given time. And even this scheme does not guarantee that some interesting tweets are not missed simply because their source has not been known to the user. It is desirable to identify which tweets are good to recommend (automatically) to the user, where these tweets have to be very close to their interests. Sources or clues to user’s interests are their twitting history and their social relations.

We propose an approach to tweets recommendation that is based on identification of other accounts that are likely sources of interesting tweets. We devised a simple scheme that is able to recommend who to follow. Experiments show our approach performs similarly or better than a known solution.

The structure of the rest of the paper begins with a brief commentary of related works. This is followed by our explanation of our approach to recommendation of tweet sources. Then we give evaluation and results; and finally, conclusions and future work are presented.

2 RELATED WORKS

The present work falls within a broader context from several points of view. Connection between personalisation and user modelling has been intensively studied by many scholars (e.g. Barla 2011) and particularly in social media (Yin 2015). The role of a group of users in personalised recommendation has been stressed e.g. by Kompan (2013, 2014). Microblogs themselves are sources of valuable information and thus a subject to analysis aiming to identify opinions (Machova 2013) and sentiments (Korenek 2014) or to perform exploratory search (Zilincik 2013).

J. Chen et al. (2010) experimented with recommending content from information streams. In designing a recommender, they explored various options. They contemplated a three dimensional design space: along the first dimension, various options of how to select candidate accounts; along the second one, various options of how to use content information; and along the third one, how to use social information.

K. Chen et al. (2012) proposed a collaborative ranking model for recommending interesting tweets. The model collects preference information from many
users. This facilitates collaborative filtering that produces recommendations. Their approach is quite comprehensive. It takes into account the content of the tweet, user’s social relations and certain other explicitly defined features.

An important issue has been raised by Liu (2014) who proposed an approach to personalised tweet recommendation that aims to be privacy preserving. Their framework provides for keeping the content of tweets and users’ interests hidden from other unauthorised entities.

A tweet recommendation method would benefit from knowing user’s topics of interest. Bhattacharya et al. (2014) proposed a method to infer the topics of interest for an individual user. Their idea is to infer them from the topical expertise of the users whom the user follows.

Our idea is to exploit the social relations even further. The user is a follower of a host of other users. There are possibly many other followers of the same set of users. What do these followers have in common?

3 FELLOWSHIP BASED APPROACH

Our assumption is that on Twitter, the semantic relationship in the followership process that can be characterized as follows/is_followed is sufficiently information-rich to yield genuine recommendations of who to follow. The question is how such recommendations can be elicited from the data that represents the users and their relations, i.e. lists of followers of each particular user.

Data on each particular user (retrieved during their logging in) includes accounts that they follow. However, often a user follows dozens, if not hundreds of other accounts. To produce a realistic recommendation, this set must be substantially narrowed.

We should bear in mind that the goal is to produce personalised recommendations to the user. Therefore, any expressions of any user A are in principle possible sources of clues of their interests. In our approach, there is retrieved their timeline and extract of up to 200 most current tweets written by the user A. They are analysed to find mentions of the followed accounts in their tweets or shared tweets (retweets). Each mention or retweet increases priority of the followed account.

There are also retrieved up to 200 favourite tweets, i.e. those marked by a star by the user A, which is an attribution similar to the “like” in Facebook. Accounts who authored such tweets receive an increase of their priority.

As a result of such priority attributions, there is formed an ordered list of accounts. We take the 10 top accounts as sources of data (we consider the bottom 10 accounts to be candidates for deletion). The crucial design decision was here to determine which data will be used for creation of recommendations. It is to be noted that due to the nature of communications occurring on Twitter, the early common recommendation data model of considering a group of related individuals (Balabanović 1997) is not sufficient since that would mean considering only accounts followed by the user (i.e., their friends). Another model includes into consideration also friends of friends of the user (Facebook) or accounts followed by accounts followed by the user (Twitter) (Chen 2010). However, this model is not as suitable for Twitter as for Facebook. The reason is that users on Twitter frequently follow not only their friends, but also publicly popular individuals, be they actors, singers, sportspersons, politicians, news services, entertainers etc. Such individuals are not user’s friends.

The key point of our model is to form a group of individuals who follow one or more accounts that are also followed by the user. This is a conceptual deviation from the commonly shared view to focus on user’s friends. We start from a list of the top 10 accounts of user A (see above). The number 10 is of course only a subjectively set ad hoc parameter of our method; on the other hand, we experimented with different values. From the data on each of these accounts, there are randomly selected 100 (again ad hoc parameter) followers, all in all totalling up to 1000 users. They create user A’s fellowship of followers – FF(A).

![Figure 1. Recommendation based on fellowship of followers](image)
Users in FF(A) have at least this in common: they follow accounts that are among the top 10 accounts followed by the user A. The next step is to find out which accounts are the most followed by users from FF(A). For each account, up to 200 (yet another ad hoc parameter) followees are extracted.

We have now at most 200,000 accounts that are followed by someone from the FF(A). Due to overlapping interests of different followers, the number is often significantly smaller, say around 50,000.

We should like to emphasize that values for the above mentioned ad hoc parameters are the result of extensive careful experimenting aimed at achieving quality of recommendation in a short time.

The multiset of several tens of thousands of accounts needs to be massively reduced. Bear in mind we want to make recommendation to the user A which accounts to follow. The multiset contains (occurrences of) accounts followed by those who have similar following portfolio as the user A. The first rule of thumb is obviously to use the number of their occurrences in the multiset. However, this does not suffice. Among the accounts with most occurrences in the multiset there can be found accounts that are in no way representative of the followers’ interests but simply have an extremely high number of followers. For example, singer Katy Perry has (as of time of writing this text) more than 71 million followers. In our approach, filtering out the less occurring accounts is augmented by a mechanism of weighting numbers of occurrences by numbers of followers (details of the mechanism are described below). In such a way, too popular accounts were moved down in the list.

4 EVALUATION AND RESULTS

We developed a prototype version of a system (MA) that implements the described approach. The MA prototype was used initially to adjust several ad hoc parameters of our method with the aim of increasing the quality of recommendation while maintaining a short response time.

The actual experimenting is aimed at an evaluation of the proposed approach. In order to be able to make any judgements on the quality of our recommendations, we need to solicit evaluating feedback from the users. To achieve as much objective evaluation as possible, we arranged for presenting recommendations by two different tools, i.e. by MA prototype and by the “Who To Follow” application of Twitter (WtF). Each evaluator was presented two lists of recommendations. In the first list, there were top 10 recommendations produced by MA. In the second list, there were top 10 recommendations produced by the WtF application. The user did not know which list was produced by which recommendation. The evaluator was asked to indicate in both lists those accounts that he would be interested in and would be willing to follow. An example of the two lists presented to the user is in Figure 2.

There were 24 evaluators – all of them Twitter users. First, recommendations by MA were produced solely based on numbers of occurrences of accounts in the multiset (i.e., no weighting). Results are in Figure 3, see columns WtF and MA.

Axis x shows evaluators. Axis y shows the numbers of accounts that the user decided to follow based on the recommendation of the respective tool (either WtF or MA). It is evident that the results for both tools are very similar. The users liked on average 24.6% recommendations produced by the WtF. They liked 25.4% recommendations produced by MA.
We have not been content with these results. We identified popular accounts as the reason for not-so-good performance of MA. This observation should not lead to deleting those highly popular accounts since it still may be the case that they could be interesting to the user A. This was endorsed by two evaluators who stated that they would decide to follow such accounts even if they do not fit to their scope of interest. We developed the following weighting scheme to diminish the preference of highly followed accounts (NoO stands for number of occurrences and a is an account from the multiset).

\[
\text{if } \text{NoO}(a) > 10^7 \text{ then } \text{NoO}(a) = \frac{\text{NoO}(a)}{10} \\
\text{else if } \text{NoO}(a) > 10^6 \text{ then } \text{NoO}(a) = 10^6
\]

In the second round of evaluation, the evaluators were presented recommendations produced by MA employing the weighting scheme. Results are in Figure 3. (See columns MA*).

In this paper we studied how to improve recommendation of Twitter accounts. This, of course, has been studied by several authors, but it remains a research topic currently. Our approach is based on a simple idea. Any user on Twitter is characterized by the set of accounts they follow. We retrieve other individuals who follow at least one of the most favourite from these accounts to form a fellowship of followers. Preferences and tastes of this group determine recommendations to the user.

Experiments, albeit limited, allow us to conclude that the approach produces better recommendations.

Acknowledgements

This work was partially supported by the Scientific Grant Agency of Slovakia, grants No. VG 1/0752/14 and VG 1/0646/15.

REFERENCES


