iSCUR: Interest and Sentiment-Based Community Detection for User Recommendation on Twitter

Davide Feltoni Gurini, Fabio Gasparetti,
Alessandro Micarelli, and Giuseppe Sansonetti

Department of Engineering,
Artificial Intelligence Laboratory, Roma Tre University,
Via della Vasca Navale, 79, 00146 Rome, Italy
http://ai-lab-03.dia.uniroma3.it/people/

Abstract. The increasing popularity of social networks has encouraged a large number of significant research works on community detection and user recommendation.
The idea behind this work is that taking into account users' attitudes toward their own interests can bring benefits in performing such tasks.
In this paper we describe (i) a novel method to infer sentiment-based communities without the requirement of obtaining the whole social structure, and (ii) a community-based approach to user recommendation. We take advantage of the *SVO (sentiment-volume-objectivity)* user profiling and the weighted Tanimoto similarity to evaluate user similarity for each topic. Afterwards we employ a clustering algorithm based on modularity optimization to find densely connected users and the Adamic-Adar tie strength to finally suggest the most relevant users.
Preliminary experimental results on Twitter show the benefits of the proposed approach compared to some state-of-the-art user recommendation techniques.

Keywords: Sentiment analysis, social networks, community detection, user recommendation

1 Introduction

Recently user recommendation on Twitter[1] has gained a lot of importance mostly thanks to the growing popularity of this social network. There are also extensive works on detecting social network communities on it, especially by characterizing contents and tags using topic extraction tools. Apart from a few notable exceptions, state-of-the-art approaches for user recommendation that rely entirely on tweet content have low precision as tweet contents are typically short and noisy, while collaborative filtering approaches that utilize follower-followee relationships lead to higher precision but data sparsity remains a challenge. The

[1] twitter.com

rationale of this work is that users can have similar interests but different opinions on them. Therefore, considering the contribution of user sentiments may yield benefits in the processes of community detection and user recommendation. In this paper, we propose a novel framework - named iSCUR - that is able to extract the implicit sentiment from tweets, to exploit it for inferring communities structure, one for each topic, and to suggest similar users that belong to the same communities. Forming communities enables us to reduce data sparsity and focus on discovering the latent characteristics of communities instead of individual users. The proposed framework proves effective in improving the recommendation precision (by as much as 5%) as demonstrated by results of online and offline preliminary experiments.

The paper is structured as follows. The community detection model is described in the next section, whilst the user recommendation process is presented in Section 3. Section 4 reports the results of a comparative experimental analysis and Section 5 reviews some related works. Finally, Section 6 concludes and outlines some possible future works.

2 Topic Community Detection

The idea behind this work is that taking into account user attitudes towards his own interests can yield benefits in recommending friends to follow. Specifically, we consider (i) which is the sentiment expressed by the user for a given concept, (ii) how much he is interested in that concept, and (iii) how much he expresses objective comments on it. For *concept* we mean any entity (e.g., hashtag, topic, etc.) extracted from a tweet that can somehow characterize it.

In our model the first contribution $S(u, c)$, namely, the *sentiment* of the user u about a concept c, is obtained as follows:

$$S(u, c) = norm\left(\frac{Pos(u, c) - Neg(u, c)}{Pos(u, c) + Neg(u, c)}\right) \tag{1}$$

where $Pos(u, c)$ and $Neg(u, c)$ are the sums of the positive and negative tweets written by the user u regarding the concept c, respectively. Such values are calculated by means of a supervised Machine Learning algorithm based on a Naïve Bayes classifier. The *norm* function is used to normalize the output value within the $[0, 1]$ range:

$$norm(x) = \frac{1}{1 + (k^{-x})} \tag{2}$$

where k = 10.

The second contribution is the *volume* $V(u, c)$, that is, how much a user u wrote about a specific concept c and is defined as follows:

$$V(u, c) = \frac{tweets(u, c)}{\sum_{i=1}^{N} tweets(u, c_i)} \tag{3}$$

where $tweets(u, c)$ is the number of tweets written by the user u about a specific concept c, and N is the total number of concepts dealt with by u.

The third contribution is the *objectivity* $O(u,c)$, which expresses how many tweets about a concept c are objective, namely, do not contain sentiments or opinions. $O(u,c)$ is defined as follows:

$$O(u,c) = \frac{Neutral(u,c)}{Pos(u,c) + Neg(u,c) + Neutral(u,c)} \tag{4}$$

where $Pos(u,c)$, $Neg(u,c)$ and $Neutral(u,c)$ are the sums of the positive, negative, neutral tweets written by the user u relative to the concept c, respectively.

Based on such contributions, we define a vector - called *sentiment-volume-objectivity (SVO)* vector, which takes into account all of them. If we consider a user u_i and a concept c, it is defined as follows:

$$\boldsymbol{SVO(u_i, c)} = [\alpha S(u_i,c), \beta V(u_i,c), \gamma O(u_i,c)] \tag{5}$$

where α, β, and γ are three constants $\in [0,1]$, such that $\alpha + \beta + \gamma = 1$. In order to determine the optimal values of those parameters, we implemented a *mini-batch gradient descent* algorithm. The best user recommender performance was achieved with $\alpha = 0.3$, $\beta = 0.6$, and $\gamma = 0.1$. Hence, based on the proposed model and the used dataset, these weights appear to favor the contribution of the *volume* and the *sentiment* in comparison with the *objectivity*.

For each concept c we compute the Tanimoto similarity [7] between users u_i and u_j as follows:

$$sim(u_i, u_j, c) = \frac{\boldsymbol{SVO(u_i, c)} \cdot \boldsymbol{SVO(u_j, c)}}{\parallel \boldsymbol{SVO(u_i, c)} \parallel^2 + \parallel \boldsymbol{SVO(u_j, c)} \parallel^2 - \boldsymbol{SVO(u_i, c)} \cdot \boldsymbol{SVO(u_j, c)}} \tag{6}$$

The similarity value lies in between $[0,1]$.

Once the similarities between users are computed, we build a graph for each concept as follows. If the similarity value between users exceeds a threshold value, we consider an edge between them. Also the optimal threshold value was determined through a gradient descent algorithm that maximizes the recommender precision. Such value was 0.8. Afterwards a clustering algorithm based on modularity optimization [4] allows us to detect the user communities for the considered concept c (see Fig. 1).

3 Recommender System

In our recommendation system, a generic user u is profiled as follows:

$$p(u) = \{(c, \boldsymbol{SVO(u, c)}) | c \in C_u\} \tag{7}$$

where the vector $\boldsymbol{SVO(u, c)}$ gives the relevance of the concept c for the user u, and C_u is the set of concepts cited by the user u. The user profile representation is generated by monitoring the user activity, that is, all the tweets included in the observation period.

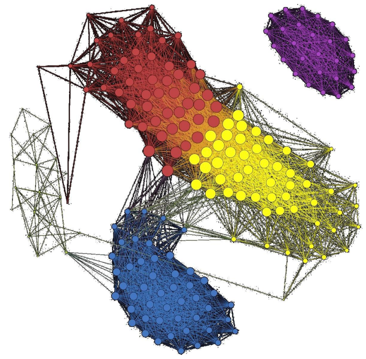

Fig. 1. Graph layout with the communities detected for one concept c.

Once we have identified the communities for all the concepts mentioned by the user u_i we want to suggest other users to follow, the recommender system works as follows. For every user u_j in the dataset, for each concept c mentioned by him we verify if it was also mentioned by the user u_i. In the negative case we analyze the next concept, in the positive case we consider the related graph. If the users u_i and u_j are connected by an edge, namely they belong to the same community, we consider the weight of the edge, which is equal to their Tanimoto similarity $sim(u_i, u_j, c)$. If there is no edge between the two users, we calculate the measure of tie strength according to the metric proposed by Adamic and Adar [2]:

$$TieStrength(u_i, u_j) = \sum_{P \in \Gamma(u_i) \cap \Gamma(u_j)} \frac{1}{log|P|} \tag{8}$$

where $\Gamma(u_i)$ and $\Gamma(u_j)$ are the neighborhoods of the users u_i and u_j respectively, and P is the number of nodes belonging to both of them. To calculate the total score between the two users, we consider the sum of all the previous contributions:

$$Score(u_i, u_j) = \sum_{c \in C_{u_i} \cap C_{u_j}} s(u_i, u_j, c) \tag{9}$$

with

$$s(u_i, u_j, c) = \begin{cases} sim(u_i, u_j), & \text{if } \exists\, edge(u_i, u_j) \\ TieStrength(u_i, u_j), & \text{otherwise} \end{cases} \tag{10}$$

We calculate the total score between the user u_i and all the users u_j and we suggest to him the N ranked users for which this value is highest. The pseudocode of the entire algorithm follows. We suppose that user u_i is the user we want to suggest someone relevant to follow.

4 Experimental Evaluation

In order to evaluate the proposed model, we considered the 2013 Italian political elections. Using the Twitter APIs we selected 31 hashtags and keywords for

Algorithm 1: iSCUR algorithm

```
input  : social timeline of user input u_i
output: top-n recommended users to follow
profile concepts from user timeline: chapter 3;
foreach concept c extracted from user u do
    foreach user u_j in the evaluation set do
        (/* calculate similarity values                                  */);
        sim(u_i, u_j) ← SVOTanimotoSimilarity;
        (/* discard users with low similarity                            */);
        if (sim(u_i, u_j) > θ) then
            (/* graph related to concept c                               */);
            Graph(c, nodes) ← insert(u_i, u_j) ;
            Relationship(u_i, u_j) ← sim(u_i, u_j);
    (/* identify communities for concept c graph                         */);
    Communities ← ClusteringAlgorithm(Graph);
(/* recommender system algorithm                                         */);
foreach graph g(c) related to a concept c do
    (/* calculate ranking score for (u_i, u_j)                           */);
    if (u_i, u_j) are in the same community then
        Score(u_i, u_j) ← ∑_{graph(c)} sim(u_i, u_j);
    else
        Score(u_i, u_j) ← ∑_{graph(c)} TieStrength(u_i, u_j);
(/* Select top-n best ranked users                                       */);
```

retrieving the Twitter streams about politician leaders and parties from Jan 25th to Feb 27th. The final dataset counted 1085000 tweets. We selected 1000 users that (i) posted at least 10 tweets in the observed period, and (ii) had more than 15 friends and followers already stored into the dataset.

The goal of our recommender is to suggest to a user someone to follow. In order to compare different profiling approaches and recommendation strategies, we need to understand when a user u_i is relevant for a user u_j. In this work we suppose that u_i is relevant for u_j if a *following relationship* exists between them. This assumption has already been proposed in literature [1, 5, 3] and is supported by the phenomenon of *homophily*, that is, the tendency of individuals with similar characteristics to associate with each other.

We performed a preliminary evaluation in order to assess the effectiveness of the proposed approach. For the sake of brevity, in this paper we only report the results of a comparative analysis of our approach with two traditional approaches that do not consider sentiment: (i) cosine similarity in a Vector Space Model (VSM) where vectors are weighted hashtags or topics extracted through WikipediaMiner [2], and (ii) the function $S1$ proposed by Hannon et al. [6]. We used the *Success at Rank K (S@K)* metric, which provides the mean probability that a relevant user is located in the top K positions of the list of suggested users. Figure 2 shows the obtained evaluation results. As can be seen, our approach outperforms the other ones. These findings confirm that sentiment is a valuable feature to be considered in order to improve user recommender systems. Generally speaking, in our tests however we have noticed the need for a deeper analysis

[2] wikipedia-miner.cms.waikato.ac.nz/

to understand the real interactions between communities, a simple calculation of tie strength is not enough.

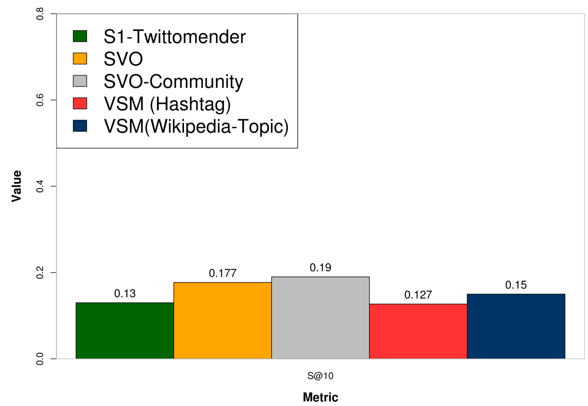

Fig. 2. Comparative analysis among the proposed approach and two other state-of-the-art methods.

5 Related Work

In spite of the growing body of research on exploiting user-generated contents in recommendation engines, there are few attempts to consider sentiment included in micro-posts for community detection or user recommendation. In [8] the authors propose ...

Twittomender [6] lets users find pertinent profiles on Twitter exploiting different strategies, both content-based and collaborative ones. Arru *et al.* [3] propose a signal-based representation of user interests in order to draw similarities among people.

6 Conclusion

In this paper we have described an approach to community detection for user recommendation. Our work emphasizes the use of implicit sentiment analysis in order to improve the recommender performance. We have defined a novel weighting function that takes into account sentiment, volume, and objectivity related to the user interests. This technique allowed us to build more complete user profiles than traditional content-based approaches. Preliminary results show the benefits of our proposed model compared with some state-of-the-art methods.

As future work, we plan to take into accounts other elements (e.g., named-entities, persons, products) and semantic representations of hashtags. A future study will also focus on the use of the implicit sentiment analysis within the collaborative filtering in social networks.

References

1. Abel, F., Gao, Q., Houben, G.J., Tao, K.: Analyzing temporal dynamics in twitter profiles for personalized recommendations in the social web. In: Proceedings of ACM WebSci '11, 3rd International Conference on Web Science, Koblenz, Germany. ACM (June 2011)
2. Adamic, L.A., Adar, E.: Friends and neighbors on the web. Social Networks 25(3), 211–230 (2003)
3. Arru, G., Feltoni Gurini, D., Gasparetti, F., Micarelli, A., Sansonetti, G.: Signal-based user recommendation on twitter. Social Recommender Systems 2013 (2013)
4. Blondel, V., Guillaume, J., Lambiotte, R., Mech, E.: Fast unfolding of communities in large networks. J. Stat. Mech p. P10008 (2008)
5. 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. pp. 199–206. RecSys '10, ACM, New York, NY, USA (2010)
6. Hannon, J., McCarthy, K., Smyth, B.: Finding useful users on twitter: twittomender the followee recommender. In: Proceedings of the 33rd European Conference on Advances in Information Retrieval. pp. 784–787. ECIR'11, Springer-Verlag, Berlin, Heidelberg (2011)
7. Tanimoto, T.: Internal report. Tech. rep., IBM (1957)
8. Xu, K., Li, J., Liao, S.S.: Sentiment community detection in social networks. In: Proceedings of the 2011 iConference. pp. 804–805. iConference '11, ACM, New York, NY, USA (2011), http://doi.acm.org/10.1145/1940761.1940913