Analysis of Sentiment Communities in Online Networks

Davide Feltoni Gurini, Fabio Gasparetti,
Alessandro Micarelli and Giuseppe Sansonetti
Roma Tre University
Department of Engineering
Via della Vasca Navale 79
Rome, 00146 Italy
{feltoni,gaspare,micarel,gsansone}@dia.uniroma3.it

ABSTRACT

This article reports our experience in developing a recommender system (RS) able to suggest relevant people to the target user. Such a RS relies on a user profile represented as a set of weighted concepts related to the user's interests. The weighting function, we named *sentiment-volume-objectivity (SVO) function*, takes into account not only the user's sentiment toward his/her interests, but also the volume and objectivity of related contents. A clustering technique based on modularity optimization enables us to identify the latent sentiment communities. A preliminary experimental evaluation on real-world datasets from Twitter shows the benefits of the proposed approach and allows us to make some considerations about the detected communities.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: [Information Filtering]

General Terms

Algorithms, Experimentation

1. BACKGROUND

The rationale behind this work is that users may share similar interests, but have different opinions about them. Therefore, we extend the traditional approaches to user recommendation through sentiments and opinions extracted from the user-generated content in order to improve the accuracy of suggestions. This way, we can identify latent *sentiment communities* of users. As far as we know, there are few works on considering users' attitudes for community detection or user recommendation. In [6] the authors formulate the problem of sentiment community discovery as a semidefinite programming (SDP) problem and solve it through an SDP-based rounding method. Nguyen *et al.* [5] address the problem of clustering blog communities into groups based on users' sentiments, and propose a non-parametric clustering algorithm for its solution. Yang and Manandhar [7] propose two community discovery models by combining social links, author based topics and sentiment information to detect communities with different sentiment-topic distributions. Unlike previous works, we consider not only the target user's attitudes toward his/her interests, but also the volume and objectivity of related generated contents.

2. PROPOSED APPROACH

Traditional approaches to user recommendation rely on the definition of a similarity measure between two users u and v. Given the target user u, the ranked list of suggested users corresponds to the set of users v that maximize the aforementioned measure. Content-based approaches on Twitter [1] define this measure by analyzing user tweets. Concepts dealt with by a user are identified through *hashtags* contained in his/her tweets, namely, the metadata tags that are used in Twitter to indicate the context or the flow a tweet is associated with. Thus, we define the profile p of the user u as the set of weighted concepts:

$$p(u) = \{(c, \omega(u,c)) | c \in C_u\} \tag{1}$$

where $\omega(u,c)$ is 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. Afterwards, given two users u and u, and their profiles $p(u)$ and $p(v)$, the similarity function is defined in terms of cosine similarity:

$$sim(u,v)=sim(p(u),p(v))=\frac{\sum_{c\in C_u \cup C_v} \omega(u,c)\cdot\omega(v,c)}{\sqrt{\sum_{c\in C_u} \omega(u,c)^2}\cdot\sqrt{\sum_{c\in C_v} \omega(v,c)^2}} \tag{2}$$

where C_u and C_v are the concepts in the profiles of users u and v, respectively. The idea behind this work is that taking into account users' attitudes towards their interests can yield benefits in recommending friends to follow. Specifically, we consider three contributions: 1) $S(u,c)$, that is, the sentiment expressed by the user u for the concept c; 2) $V(u,c)$, that is, how much he/she is interested in that concept; 3) $O(u,c)$, that is, how much he/she expresses objective comments on it. The details regarding the computation of such contributions can be found in [2]. Based on those contributions, we propose a weighting function, we called *sentiment-volume-objectivity (SVO)* function, that takes into account

[1] `twitter.com`

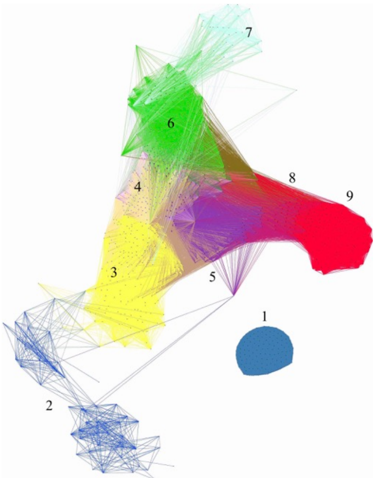

Figure 1: Communities for the *Apple* concept, detected through the SVO profiling and clustering.

all of them. It is defined as follows:

$$SVO(u, c) = \alpha S(u, c) + \beta V(u, c) + \gamma O(u, c) \qquad (3)$$

where α, β, and γ are three constants $\in [0, 1]$, such that $\alpha + \beta + \gamma = 1$. The function $SVO(u, c) \in [0, 1]$ is the weighting function $\omega(u, c)$ that appears in Equations 1 and 2. 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 θ, we consider an edge between them. The optimal value for θ 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 [1] allows us to detect the latent communities for the considered concept c.

3. EXPERIMENTAL EVALUATION

Experimental tests were performed on three datasets gathered from Twitter through its APIs [2] by searching for specific hashtags. $Dataset_1$ was obtained during the 2013 Italian political elections. We retrieved the Twitter streams about politician leaders and Italian parties from January 25th to February 27th. The final dataset counted 1,085,121 tweets in Italian language and 70,977 unique users. $Dataset_2$ was obtained searching for hashtags and keywords representing the most important mobile tech companies such as Samsung, Apple, and Nokia. The dataset was gathered from September 2014 to February 2015 considering only Italian tweets, and counted 3,511,455 tweets from 181,000 users. $Dataset_3$ was obtained analyzing English tweets on the automotive landscape. To this aim, we searched terms such as Audi, BMW, and Ferrari. The collection set, gathered from December 2014 to February 2015, counted 2,915,131 tweets from 110,350 users. Figure 1 shows the different communities detected for the Apple concept ($dataset_2$). The bottom, right, isolated, community marked with number one includes users not interested in Apple ((i.e., their SVO value is zero). The bottom, left, community with number two consists of users with low interest (i.e., low value of volume) and negative sentiment. The community labeled with number three is characterized by high interest and negative sentiment. The central communities identified with numbers four, five, six, and seven encompass users with high interest and positive sentiment (with different SVO combinations for each community). Users belonging to communities eight and nine have high values of objectivity, namely, they generate objective contents with few opinions. Among such users, for example, we find online newspapers such as BBC and CNN.

[2] `dev.twitter.com`

Table 1: A comparison among different techniques

User Recommender	$Dataset_1$	$Dataset_2$	$Dataset_3$
Our Approach	**0.177**	**0.195**	**0.185**
S1-Twittomender	0.130	0.118	0.115
VSM (Hashtag)	0.127	0.099	0.105

Table 2: SVO parameters for the three datasets

Parameter	$Dataset_1$	$Dataset_2$	$Dataset_3$
α	0.45	0.25	0.28
β	0.45	0.50	0.52
γ	0.10	0.25	0.20

In order to evaluate our approach, we relied on the *homophily* [4] phenomenon, that is, the tendency of individuals with similar characteristics to associate with each other. For each dataset, we selected 1000 users that (i) posted at least 50 tweets in the observed period, and (ii) had more than 30 friends and followers. Table 1 reports the results in terms of *Success at Rank 10 (S@10)* of a comparative analysis of our system with two state-of-the-art functions: 1) a content-based function, called *S1-Twittomender* [3], where users are profiled through the content of their tweets; and 2) a *VSM (Hashtag)* function representing cosine similarity in a vector space model, where vectors are weighted hashtags. Table 2 shows the values of SVO parameters α, β, and γ that maximize the performance of the recommender. Such values were determined through a mini-batch gradient descent algorithm. Based on the proposed model and the used datasets, these weights appear to highlight the contribution of volume and sentiment in $dataset_1$, and objectivity in $dataset_2$ and $dataset_3$. This can be explained because $dataset_2$ (technology) and $dataset_3$ (automotive) are likely to contain more news and articles with few opinions and sentiments than $dataset_1$ (politics).

4. CONCLUSIONS AND FUTURE WORK

In this article, we have presented some results of our research work whose aim is to exploit the implicit sentiment analysis for improving the performance of user recommenders. In particular, we have reported some preliminary considerations on the sentiment communities that our approach allows us to identify.

Our work is still at an exploratory stage, so several its aspects have to be further developed. Among others, we plan to perform an in-depth sensitivity analysis to study how user preferences, social interactions, and dataset characteristics can affect SVO parameters α, β, and γ. Their role is indeed crucial, since they define the contribution extent of sentiment, volume, and objectivity that determine the distribution of users in sentiment communities.

5. REFERENCES

[1] V. Blondel, J. Guillaume, R. Lambiotte, and E. Mech. Fast unfolding of communities in large networks. *J. Stat. Mech*, 2008.

[2] D. F. Gurini, F. Gasparetti, A. Micarelli, and G. Sansonetti. A sentiment-based approach to twitter user recommendation. In *Proc. of the Fifth RSWeb Workshop co-located with the Seventh ACM RecSys Conference, Hong Kong, China*, 2013.

[3] J. Hannon, M. Bennett, and B. Smyth. Recommending twitter users to follow using content and collaborative filtering approaches. *Proceedings of the 4th ACM RecSys Conference*, 26-30(10):8, 2010.

[4] M. McPherson, L. S. Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444, 2001.

[5] T. Nguyen, D. Q. Phung, B. Adams, and S. Venkatesh. A sentiment-aware approach to community formation in social media. In *ICWSM*. The AAAI Press, 2012.

[6] K. Xu, J. Li, and S. S. Liao. Sentiment community detection in social networks. In *Proc. of the 2011 iConference*, pages 804–805, New York, NY, USA, 2011.

[7] B. Yang and S. Manandhar. Community discovery using social links and author-based sentiment topics. In *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on*, pages 580–587, 2014.