Sentiment-Based Influence Detection on Twitter

Carolina Bigonha[1], **Mirella M. Moro** ***(advisor)***[1], **Marcos A. Gonçalves** ***(co-advisor)***[1]

[1]Departamento de Ciência da Computação, UFMG, Belo Horizonte, MG, Brazil
{carolb,mirella,mgoncalv}@dcc.ufmg.br

Abstract. *We propose a method for answering the question on how to find influential users for a topic in large online communities. This method for ranking users in Twitter is based on a combination of the users' position in networks that emerge from their relations, the polarity and the textual characteristics of their posts. Our evaluation shows that our approach can successfully identify influential users on different datasets.*

Resumo. *Nessa dissertação, apresenta-se um método para identificar usuários influentes para tópicos no Twitter, que se baseia em aspectos comportamentais do usuário: sua posição em redes de conversação, sua polaridade e o conteúdo de seus tweets. Nossa avaliação experimental demonstra êxito em identificar os usuários mais influentes em diferentes bases de dados.*

1. Introduction

Twitter is a widely used micro-blogging tool that represents a real-time information network. Users of Twitter share opinions and experiences on tweets of up to 140 characters. Considering Twitter users as potential consumers and the Word of Mouth (WOM) generated by their discussions, micro-blogging networks have become a rich source of data in any situation in which feedback is desired [Brown et al., 2007]. By studying the data and the users, businesses can gather market intelligence and improve their campaigns, products or services acceptance. Analyzing this data is not simple, though, due to the huge amount of content generated daily. Besides being impractical to inspect all the data, even for a specific topic, not all tweets and users are worth such an evaluation. Under these circumstances, in order to save time (and resources), it is crucial to find the opinion leaders, or *influential users*, who drive WOM conversations on Twitter. Katz et al. [1955] defined as *opinion leaders* "the individuals who were likely to influence other persons in their immediate environment". By identifying these key users, marketers can benefit from a social multiplier effect on their marketing efforts and leverage lower (and strategic) investments.

In this work, we explore *sentiment-based influence given a topic*. Our focus is on topics due to the usual interest in monitoring one particular context, e.g., products, personalities, events. And it is *sentiment-based* motivated by insights that can be extracted from positive and, especially, negatively biased content. The intuition is that negative posts are more likely to induce consumers to change their mind about a product than positive ones [Lee et al., 2008]. Thus, identifying negatively biased users may simplify the marketing analysis for branding strategy and brand-customer interaction.

Our method, called SaID (Sentiment-Based Influence Detection on Twitter), focuses on uncovering the behaviour of users based on Twitter data and telling influential and not influential ones apart. The main contributions of this work are: *(1)* a new and clear definition of what an influential user is; *(2)* the SaID method for detecting influential users based on the polarity of their tweets; *(3)* fully analyzed datasets that are available online [1]

[1]**Dataset available for download:** http://goo.gl/tExDj

and can be used as benchmark for future work; *(4)* detailed comparison of the effectiveness of interactions via tweets and following connections on determining influence; *(5)* a thorough discussion on how polarity, relation and content may affect influential detection; *(6)* considerations about the effect of automatic tweet classification on influence detection.

In addition to the aforestated contributions, this work has produced two publications: [Bigonha et al., 2010], awarded best paper of WebMedia and already cited by 16 other papers, and [Bigonha et al., 2012], which is its extended journal version.

2. Sentiment-Based Influence Detection

Focusing on a marketing and consumer point of view [Kwon and Sung, 2011] an **Influential User** is the one *(i)* whose actions imply in other persons' actions; *(ii)* who acts like bridges on interactions about a subject; *(iii)* who has a positive or negative bias on their opinion; *(iv)* who produces content with a minimum expected quality.

The first Item is directly derived from the basic definition of influence and opinion leaders: it focuses on the fact that influential users' actions cause effect on others. Meanwhile, Item *ii* evaluates the centrality of the users in the discussions. In order to maintain a leadership role on a topic, the user has to be a part of the active discussion: generating buzz around their posts and acting like bridges on interactions. Next, motivated by the insights can be extracted from polarized content, Item *iii* implies that, in order to influence others, a user has to have a positive or negative bias in their opinion. Finally, by *minimum quality*, in Item *iv*, we mean well structured sentences, with the intention of presenting an idea. Influential users are not occasionally talking about the topic, they have a purpose for posting content. SaID is a method based on the aforementioned characteristics and is divided in three phases: *pre-processing*, *feature extraction* and *influence score*.

2.1. Pre-processing

The first step in the pre-processing phase is to determine *which topic is going to be analyzed* and *for how long*. Based on the chosen topic, keyword-based queries are built. Also, a time interval is set because SaID calculates the user Influence Score based on a snapshot of the conversations for a topic. For gathering the content, we have built a crawling module, that uses Twitter Search API for collecting publicly available tweets, which contains the defined keywords. Looking at the collected tweets, we carefully eliminate occasional spams and tweets that fit into the keyword search, but in a different context. This process is conducted manually. After this filtering, the remaining tweets are stored. Finally, we store the authors' name, their profile URL and their list of followers and following users. We retrieve this information also using the Twitter API.

2.2. Feature extraction

For mapping users' characteristics to extractable data, three main perspectives are tackled: relation, polarity and content, as follows.

Relation perspective. We assume that the level of influence of users is directly associated with their social relation with other users in the same topic context. From the several networks that naturally emerge from user relations enabled by Twitter features, we select two of them for an in-depth analysis: the Connection Graph (G_c) and the Interaction Graph (G_i). Formally, the networks are defined as follows.

Definition 1 ***Connection Graph.*** *For a given subset of users involved in a specific topic, let (G_c, U) be the user directed unweighted graph, where (u_1, u_2) is a directed arc in U if user $u_1 \in G_c$ follows user $u_2 \in G_c$.*

Definition 2 ***Interaction Graph.*** *For a given subset of users involved in a specific topic, let (G_i, U) be the user directed unweighted graph, where (u_1, u_2) is a directed arc in U if user $u_1 \in G_i$ has cited at least once (i.e., mention, reply or re-tweet) user $u_2 \in G_i$.*

We employ centrality measures to evaluate the notoriety of users according to their position in the network. We analyze two centrality measures: **Betweenness centrality** (bc) and **Eigenvector centrality** (ec)[2]. We also analyze the **In-degree** (id) and the ratio of followers to followees of an user, tff. The most influential user for a topic is the one with the higher value for each of these metrics (tff, bc, ec, id). For this reason, the metrics were combined in an arithmetic mean and individually normalized to a $[0, 1]$ scale.

Polarity perspective. The sentiment analysis of the content allows detecting the engagement of the users towards the defined topic. Consequently, it leads to identifying well connected users responsible for influencing others' decisions due to the polarity of their tweets. We classify the tweets as positive, neutral and negative. Based on the classification of tweets, we calculate the polarity of the users, i.e., their *overall* contribution to the topic discussion. If users post mostly positive-biased content, they are considered as potential *evangelists*. On the other hand, if they post mostly negative-biased content, they may be potential *detractors*. Users that do not have a biased content are considered neutral. We consider that positive and negative tweets nullify each other. Thus, for each user, her polarity value is the summation of the sentiment of all of her tweets. Positive and negative values were range normalized separately: positive values to $[0, 1]$ and negative values to $[-1, 0]$. The normalization was calculated using logarithmic quantities.

Content perspective. Finally, we study content features of the user. We hypothesize that if users are to influence other people, their tweets are expected to have a minimum quality. As shown by Brown et al. [2007], consumers seem to evaluate the credibility of online WOM information in relation to the individual contributor of that information. For that matter, each tweet is evaluated using the Flesch-Kincaid Grade Level metric [Ressler, 1993] ($kincaid$), which was designed to indicate comprehension difficulty when reading a passage of contemporary academic English. For each tweet, it computes the average number of syllables per word and the average sentence length[3]. For instance, a tweet like "aaaaaaa haaate justin bieber!" has a low quality value, whereas "PayPal is dangerously easy." a high one. Even though the readability metrics are not expected to work flawlessly for the short sized and noisy content of tweets, the results show that the metric helps eliminating undesirable content. The user quality perspective was determined as the average of the Kincaid metric computed for each tweet of the user.

2.3. Influence score

At last, SaID combines the three perspectives into a single influence score. By exploiting them together, we aim to assign a single value (**influence score**) to each user in order to obtain a final and possibly better user rank. The user influence score is given by $I_{score} = (\alpha * u_{polarity} + \varphi * (\beta * u_{relation} + \gamma * u_{content}))/(\alpha + \beta + \gamma)$, which is one of the main

[2]Metrics calculated using **NetworkX** - https://networkx.lanl.gov/

[3]Metrics calculated using **Style and Diction Package:** http://www.gnu.org/software/diction/diction.html

contributions of this work. The variables $u_{polarity}$, $u_{relation}$ and $u_{content}$ are the normalized perspectives; α, β and γ are constants, greater or equal to zero, that weight each of the three perspectives; and $\varphi = \frac{u_{polarity}}{|u_{polarity}|}$. The auxiliary variable φ adjusts both relation and content perspectives according to the polarity result.

A feature alone may not be enough to characterize whether a user is influent or not, whereas the combination of the features may be. A user who is well positioned in the graph, has a biased opinion, and writes fairly well written tweets should be ranked higher as an influential user. The formula eliminates types of profiles that are erroneously appointed as influent. For example: *(a)* someone that is well connected to other users, but does not have a biased opinion about the subject; *(b)* someone that posts, daily, hundreds of positive/negative tweets about the topic, but, for some reason, no one pays any attention to; *(c)* a person whose content is too noisy and does not have a persuasive speech. For the specific cases listed above, the low values of polarity (a), relation (b) and content (c), respectively, do keep those users from being considered as influent.

3. Experiments

Next, we summarize the main results of this work, ommiting details due to lack of space.

Datasets. For the experiments, we have built three collections concerning conversations about brands: the first one is about *soda*, with $6,885$ users and $8,063$ tweets posted in one month; the second about *appliance*, with $1,617$ users and $2,354$ tweets also posted in one month; and the third about a *groceries megastore chain*, with $4,372$ users and $9,383$ tweets posted during a week. Even though all three datasets consist on conversations concerning a product for marketing purposes the datasets are very different from each other in terms of volume of content, number of active users, and type of posts. This dissimilarity enriches our analysis of influence. All tweets were manually classified as positive, negative or neutral by a team of specialists in marketing, who also identified the influential users for the first two datasets. This group usually provide this kind of service professionally. For the groceries megastore dataset, we determined the influential users based on a user study, in which non-specialist participants evaluated the profiles in a pool. This contrast of evaluation is directly reflected in our results. The type of user selected as influential by specialists was different from the ones selected by non-specialists. The later group was very influenced by the number of followers of the profiles in the collections.

Evaluation. In order to evaluate our method, we employ ranking performance measures: *precision*, *recall* and $\mathcal{F}_\beta$ (F-score). We evaluate each measure value at a user ranked list of size x. We use the notations $recall\,@\,x$, $precision\,@\,x$ and $\mathcal{F}_\beta\,@\,x$, considering $10 \leq x \leq 150$. Also, we consider for evaluation an approximation of the area below the measure curve, for which we use the notation $a([measure])$, e.g., $a(recall)$. We have constructed four different baselines for evaluating SaID. In Klout Baseline **(KB)** we divide the users by polarity and order them by their *Klout Score*[4]; in Tweet Baseline (**TB**) we order them by the number of tweets posted on the topic and in the Follower Baseline (**FB**) we rank the users according to their number of followers. We also employ a Random Baseline (**RB**) in which two *random* lists of users are generated: one for positive and one for negative users.

Interaction vs. Connection. Firstly, we employed a comparative analysis between two proposed approaches for finding influential users: using G_c and using G_i. We compared

[4]**Klout** http://klout.com/

the alternatives employing *paired observations*. We also analyzed the computational complexity for extracting the features bc, in and ec in G_i and G_c, on each dataset. Considering the cost of building G_c (collecting the followers of each user may be slow and expensive due to Twitter API access limits), the size of the graphs (G_i is much sparser, since users interact with way less users than they follow) and the approaches' performance when finding the influential users, we conclude that the Interaction Graph is better for determining user influence. It provides equal (groceries megastore) or even better (soda and appliance) results than the Connection approach and it is simpler to construct.

Perspective Impact. In the next experiment set, our goal was to test the hypothesis that a single perspective may not be good enough to classify users as influential or not. Besides comparing rankings generated using only one component of the Influence Score formula at a time with the baselines, we conducted a 2^k *experimental (or factorial) design*. We have concluded that the polarity perspective was the main factor for the first two datasets, while the relation perspective was rather decisive for the groceries megastore dataset. We took a deeper look at the datasets and found that the sentiment of the tweets was not determinative when detecting influentials on the last dataset: the evaluation pool participants' had a bias towards users with higher number of followers.

Parameter Estimation. Determining the combination of α, β, and γ that provides the best result is an important issue. We analyzed the combination of parameters and its effect on $a(recall)$ for detractors, by presenting ternary plots for the three parameters. By analyzing the plots, we concluded that although the values of $a(recall)$ are different between the datasets, considering the datasets individually, the result does not change much for the possible combinations of α, β and γ. Thus, by choosing an intermediary combination of the parameters, one can guarantee good results for all datasets. In this work, specifically, for estimating the potential of SaID, we optimize the parameters of the Influence Score formula. Specifically, we use a *leave one out* approach.

Evangelists x Detractors. Moreover, we discussed the final results for ranking evangelists and detractors using the Interaction-based approach. We ran *paired observations* of SaID with each baseline: we conducted 15 evaluations ($recall\,@\,x$, $10 \leq x \leq 150$) for each pair and dataset and computed the mean diference of performance in each scenario. In summary, considering all the six combinations of datasets and polarities, SaID is the best method in three scenarios (soda-evangelist, soda-detractor, appliance-evangelist), it is the best or equal in two scenarios (appliance-detractor, groceries-detractor) and worst just in one scenario (groceries-evangelists). These results clearly demonstrate the effectiveness of our method for detecting influential users.

Towards a Fully Automatic Approach. Lastly, we tackled the feasibility of a fully automatic approach, by examining the impact of an automatic sentiment analysis in the detection of the influentials. For analyzing the sentiment of the tweets, we used the state-of-art classifier `SVM` - Support Vector Machine. We performed 10-fold *cross-validation* and because the test folds did not contain intersection of instances, each tweet was classified only once. In such manner, to associate the classifier sentiment prediction with the tweets, we used the predictions for the test set, in each iteration.

Although we evaluate and discuss the results of the tweets' analysis, for detecting evangelists and detractors, the final polarity assigned to the user is more important than the accuracy of tweets classification. There is no damage in influential users detection if the overall polarity of each user is maintained. For example, the classification algorithm may predict a positive tweet as neutral and even then, the polarity of the user remain positive. When evaluating the final polarity of the users, on soda and appliance datasets at least $\sim 60\%$ of the users for each class were correctly classified, with the automatic analysis of tweets. For the groceries megastore dataset this proportion is about $\sim 70\%$.

Finally, we compare SaID results using both manual and automatic analysis of tweets by conducting *paired observations* of the experiments for all datasets. Observing the results, we note that the difference of effectiveness between the automatic and the manual approach of SaID is small. Although some users had their polarity mistaken by the automatic classification, it did not impact SaID result much. In all datasets and polarities, no more than two influentials were misclassified as not-influential, showing that the method can be fully automatized without significant effectiveness loss.

4. Concluding Remarks

This work addressed the problem of identifying biased influential users about a topic on Twitter. We presented SaID, a method that lists potential evangelists and detractors for a topic by extracting features from the users, such as the polarity and readability of their tweets and their centrality in terms of interactions via tweets. Due to the lack of space, we have only summarized the experiments conducted in our work. In terms of finding influentials, we have compared interaction and connection networks; measured the impact of each set of user characteristics; and discussed the difference between manual and automatic classification of tweets. Also, we have shown that the results produced by SaID are as good as those extracted manually. The detailed experiments and discussions may be found in [Bigonha et al., 2010], [Bigonha et al., 2012] and the dissertation itself.

Finally, we present some issues that may be addressed as future work: (1) test machine learning and rank aggregation as alternatives to rank the users; (3) improve the content perspective; (4) take into account temporal aspects when determining influence; (5) improve the tweet sentiment classification method; (6) employ a supervised (or semi-supervised) learning approach for filtering inappropriate content (pre-processing phase), in order to have a completely automatic method. The perspective of having many future work topics is further evidence of the applicability and novelty of this work.

References

C. Bigonha, T. Cardoso, M. Moro, M. Gonçalves, and V. Almeida. Detecting evangelists and detractors on twitter. In *Simpósio Brasileiro de Sistemas Multimídia e Web*, 2010.

C. Bigonha, T. Cardoso, M. Moro, M. Gonçalves, and V. Almeida. Sentiment-based influence detection on twitter. *Journal of the Brazilian Computer Society*, 18(3):169–183, 2012.

J. Brown, A. J. Broderick, and N. Lee. Word of mouth communication within online communities: Conceptualizing the online social network. *Journal of Interactive Marketing*, 21(3):2–20, 2007.

E. Katz, P. Lazarsfeld, and E. Roper. *Personal influence: the part played by people in the flow of mass communications*. Transaction Publishers, 1955.

E. S. Kwon and Y. Sung. Follow Me! Global Marketers' Twitter Use. *Journal of Interactive Advertising*, 12:4–16, 2011.

J. Lee, D. Park, and I. Han. The effect of negative online consumer reviews on product attitude: An information processing view. *Electronic Commerce Research and Applications*, 7(3), 2008.

S. Ressler. *Perspectives on electronic publishing: standards, solutions, and more*. 1993.