Behavioral Analysis in Social Networks: An Approach Based on Intelligent System

Thiago Rúbio
Computing Department
Federal Center of Tech.
Education of Minas Gerais
Belo Horizonte, MG
thmunhoz@lsi.cefetmg.br

Rogério Gomes
Computing Department
Federal Center of Tech.
Education of Minas Gerais
Belo Horizonte, MG
rogerio@lsi.cefetmg.br

Adriano Pereira
Computing Department
Federal Center of Tech.
Education of Minas Gerais
Belo Horizonte, MG
adriano@decom.cefetmg.br

ABSTRACT

This paper presents a new methodology to classify Twitter users based on Artificial Neural Networks and Fuzzy Logic. Simulations are carried out using a SOM (Self Organized Maps) neural network for classifying users into four distinct groups: (0) *Unimpressive User*; (1) *Desired User: Follower*; (2) *Desired User: Follower and Publisher*; (3) *Desired User: Publisher*. The proposed methodology was validated through an autonomous agent, whose interactions with others were modeled by means of Fuzzy Inference System. The results obtained show that neural networks can be used for user classification in social networks, and we observed that the interaction of the agent with other users occurred in a transparent way, i.e., showing typical behaviors of real users.

Categories and Subject Descriptors

I.2 [**Artificial Intelligence**]: Miscellaneous; J.4 [**Social and Behavioral Sciences**]: Miscellaneous

General Terms

Algorithms, Experimentation, Human Factors

Keywords

Social Networks, Artificial Intelligence, Machine Learning, Twitter, Self Organized Maps, Fuzzy Logic

1. INTRODUCTION

Social Networks can be seen as large systems in which one can study the behavior of users by identifying individual interests, its affinity relations, as well as their common goals [8]. The analysis of the behavior of individuals is essential for assessing the relationships among them and social groups [4, 3], enabling the creation of dynamic content management systems [1].

There are several algorithms for classifying users in social networks in the literature, most of them related to the *PageRank* algorithm [2]. Basically, this method is based on the analysis of the complex networks, as well as the counting of the interactions between users in social networks. However, these approaches may lead to a superficial understanding of the user behavior [14].

Thus, this paper presents a new approach based on SOM (Self Organized Maps) artificial neural network [13, 12] for the classification of Twitter users [3, 7, 1]. Moreover, an autonomous agent capable of interacting with others Twitter users is modeled in order to validate the proposed classification system. This agent is modeled through a fuzzy inference system and should present the same features displayed by users of a particular group.

In a summarized manner, this paper proposes the use of artificial intelligence techniques in the analysis of user behavior in social networks. Unsupervised networks are used in the process of grouping or classifying the datasets, and are also widely used as classifiers in data mining approaches. Moreover, the autonomous agent models non-numeric linguistic concepts of user behavior, enabling to design a content-based recommendation system in social networks.

2. RELATED WORK

In the virtual environment, people want to create social relations that exist or that they would like to exist in the real world. The fundamental connection between people in a virtual social network site is performed by means of identity, the establishment of friendships, as well as by means of similarity of thoughts and goals. Thus, users with similar features can be grouped and common behaviors can be used to indicate which users with similar characteristics should be allocated in the same group [2, 5]. This group, in turn, can be accomplished through the use of neural networks, which are powerful tools for classification [12].

Another important feature regarding the classification of users is the fact that the process of validating results is not a trivial task. In dynamic networks such as social networks, it becomes a challenge due to the fact that there is not an appropriate statistical model that provides the quality of the classification [10, 8].

One approach that can be used in assessing results would be the application of autonomous agents, since we can control the types of interactions among them, as well as provide them with inherent characteristics of real users [6]. Thus,

WebMedia'12, October 15–28, 2012, São Paulo/SP, Brazil.

the proposed model is based on Fuzzy Logic [9] , which allows the modeling of user behavior based on their qualitative states.

3. FUNDAMENTS AND METHODOLOGY

The aim of this paper is to evaluate the result of classification of Twitter users by means of neural networks. Therefore, we implemented a version of SOM neural network algorithm (Self Organized Maps), which consists of an unsupervised neural network training [15].

The methodology used in this article may be used in any social network, since the developed techniques are independent of the adopted technology. The choice of Twitter as a case study was a matter of design decision.

3.1 SOM Neural Network

The SOM algorithm is based on unsupervised and competitive learning. It provides a topology preserving mapping from high dimensional space to map units. Map units, or neurons, usually form a two-dimensional lattice, this it provides a mapping from high dimensional space onto a plane. The property of topology preserving means that the mapping preserves the relative distance between the points. Points that are near each other in the input space are mapped to nearby map units in the SOM. Therefore the SOM can serve as a cluster analyzing tool of high dimensional data. Moreover, the SOM has the capability to generalize. Generalization capability means that the network can recognize or characterize inputs it has never encountered before. A new input is assimilated with the map unit it is mapped to [13].

The approach adopted in this work is related to analysis of individual users. Thus, the following criteria were selected: (a) Number of Tweets; (b) Number of Followers; (c) Following Number; (d) Number of Followers/Following Number; (e) Following Number/Number of Followers; (f) Number of Replies;

These criteria are used as inputs for the neural network in order to classify users into groups. The database used in the training phase of the SOM neural network was composed of 300,000 users, obtaining at the end of the calibration a configuration with four groups, namely:

- Group 0 - *Unimpressive User*: displays low number of Tweets, Following and Followers. It has no significant relevance to influence other users.
- Group 1 - *Desired User - Follower*: it has a large number of Tweets and Following, but does not have many Followers.
- Group 2 - *Desired User - Follower and Publisher*: these users have the most interesting features, since they are very active on the network and are always sharing and seeking new information. They have high rates of Tweets, Followers and Following.
- Group 3 - *Desired User - Publisher*: with a significant number of Tweets, these users have many Followers that disseminate their new content, but therefore they have a relatively low number of following.

3.2 Fuzzy based Agent

In order to validate the classification found by the SOM neural network, it was implemented an autonomous agent whose actions and decisions were modeled using a Fuzzy Inference System. In general, a Fuzzy system is a representation that defines the qualitative states and treats them quantitatively [11, 9]. Thus, it is possible to evaluate non-quantifiable linguistic concepts, such as “very happy”, “somewhat solitary”, etc.. This model is important because it allows the creation of a user whose typical response time of the interactions with other users ensures a constant change in states of the agent.

The Fuzzy Inference System was developed having as inspiration the mood states of a human being. The mood states modeled in the agent were: Excitation, Satisfaction and Loneliness. It is important to emphasize that this choice was arbitrary, since Fuzzy Logic works, as already described, with linguistic concepts. Therefore, this is one of the biggest advantages of Fuzzy Logic and one of the main contributions of this work.

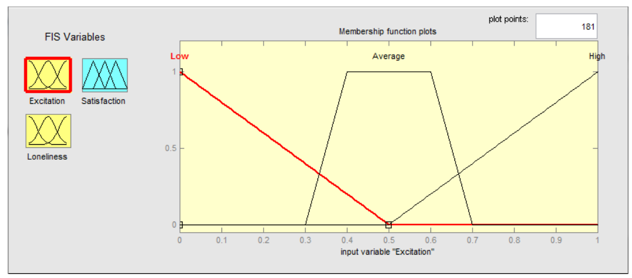

Figure 1: Fuzzy Set of the input variable Excitation

The states of Excitation and Loneliness are the linguistic variables used as input to the Fuzzy system, while Satisfaction is the linguistic variable that models the output of the system. This developed model takes into account the fact that the agent is happy when another network user interacts with him, reducing their level of Loneliness. When there are no interactions for a particular period of time, the level of Excitation decreases, followed or not by an increase in Loneliness. A critical level of Satisfaction (composed of the aggregation of the fuzzy inferences between Excitation and Loneliness) triggers an action to initiating relationships with new users (following). When the agent presents high levels of Satisfaction it generates new content, such as a comment or phrase.

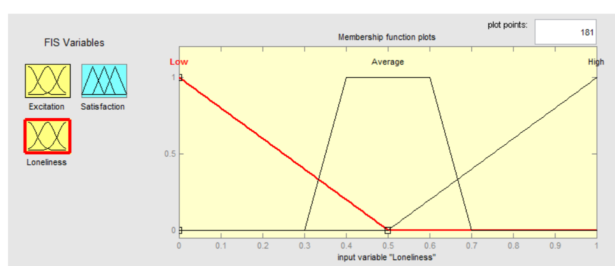

Figure 2: Fuzzy Set of the input variable Loneliness

By using the Twitter API, it was possible to integrate the agent “feelings” (implemented by the Fuzzy system) with some useful resources of the Twitter. The API works as a layer in the system, which has the responsibility to communicate with Twitter and produce the common actions of the users like posting, establishing friendship, searching etc..

The agent can call this layer as a result of the composition of the Fuzzy rules. When the output variable reaches some levels, it triggers actions in the Twitter API.

Thus, the modeled agent simulates a real user being able to perform actions, such as following users, attracting followers and posting messages (tweets). The agent implements a set of messages to post, suggest followers to other users, post comments about its mood (according to the Fuzzy state) or reference the trending topics (the most discussed terms of the network at a given time).

We identify the interest of users in the content published by the agent to the extent that users of certified accounts, such as the Official Firefox (@firefox) and others started to follow the profile of the agent. This corroborates the influence of the agent and shows that other users started to create an identifying relationship with the profile of the agent, showing interest in its content and in its friendship network.

The Fuzzy model that characterizes the agent can be seen in Figures 1, 2 and 3.

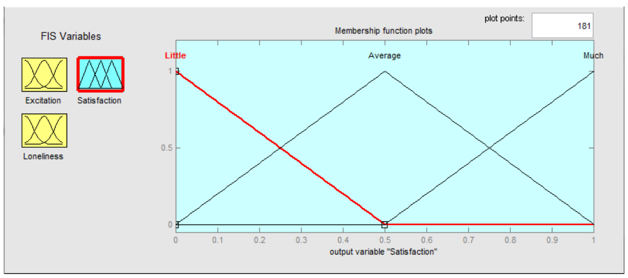

Figure 3: Fuzzy Set of the output variable Satisfaction

The fuzzy rules that implement the actions of the agent can be seen in the Code 1.

Code 1 Fuzzy model rules

```
METHOD: COG;
DEFAULT: = 0;
END_DEFUZZIFY

RULEBLOCK No1
AND: MIN / / Using 'min' for 'and'
ACT: MIN / / Using 'min' activation method
ACCU: MAX / / Using 'max' accumulation method

RULE 1: IF Excitation IS high
THEN Satisfaction IS much;

RULE 2: IF Excitation IS average
AND Loneliness IS average
THEN Satisfaction IS average;

RULE 3: IF Loneliness IS low
THEN Satisfaction IS much;

RULE 4: IF Excitation IS low
OR Loneliness IS high
THEN Satisfaction IS little;

RULE 5: IF Excitation IS high
OR Loneliness IS low
THEN happy IS very;
```

4. CASE STUDY AND EXPERIMENTS

In order to demonstrate the effectiveness of the methodology applied to Twitter, several experiments were performed to validate the system. Initially, the SOM neural network was calibrated using a database composed of 300,000 users, as described in Subsection 3.1.

After the calibration process of the SOM neural network using the training database, a system for collecting more users was developed in order to validate the classification capacity of the network. At the end, a total of 3,000 users were collected and classified. Table 1 presents a sample of 10 users classified by the neural network, highlighting the results of the classification.

It is possible to notice that the results in Table 1 are coherent with the proposed model, since users with low rates of Tweets, Followers and Following were classified in Group 0. Users with great tendency of following people content were classified, on the other hand, in Group 1. Similarly, users who are basically followed were classified in Group 3.

Table 2 shows the final result of the user classification, and it is possible to note that most users were classified as *Unimpressive User*. It is verified that this occurs because many social network users are note very participative on a global level, although they can be very influential in their circle of friendship.

Table 1: Sample data and its respective classification

Followers	Following	Tweets	Group
3	27	154	0
35	37	252	0
57	74	16	0
160	780	157	1
97	1997	602	1
1158	1641	293	1
586	268	57	1
21900	3858	5987	2
485	555	171	3
3878	1343	1057	3
1883	862	885	3

Table 2: Percentage of users rated in each group

Group	Number of Users	Percentage
0 - *Unexpressive*	1578	0.53
1 - *Follower*	846	0.28
2 - *Follower and Publisher*	109	0.04
3 - *Publisher*	467	0.16

Experiments were conducted during one week in order to promote the sort of agent to Group 2. Thus, starting from the relationship with just one user, the agent was initially classified as Group 0 by the SOM neural network (*Unimpressive User*). From that moment, the agent started the process of interaction with Twitter users, publishing content dynamically, creating following relations and seeking for followers. Every day after the interaction with other Twitter users, the agent was again classified by the SOM neural network in order to verify its performance. At this moment, we recorded the profile information of the agent, as summarized in Table 3. You can verify that in this period the agent's user passed from *Unimpressive User* (Group 0) to *Follower User*(Group 1), being finally classified as *Follower and Publisher User* (Group 2). The agent was not classified as a *Publisher User* (Group 3) due to the nature of interactions with Twitter users. The rules that trigger actions to publish massive content (tweets) were not activated and therefore the number of tweets was not very large, not allowing it to be classified as belonging to Group 3.

Figure 4 shows that there is a strong correlation between the parameters selected for the user classification, i.e., as soon as the values become more expressive it was verified that the classification performed by the SOM neural network converged to the expected results, according to the proposed model.

Table 3: Agent properties during the experiment

Followers	Following	Tweets	Date	Group
1	1	0	Day 00	0
12	27	69	Day 01	0
83	112	94	Day 02	0
192	547	448	Day 03	1
240	631	561	Day 04	1
356	712	739	Day 05	1
781	842	957	Day 06	1
1230	1035	1346	Day 07	2

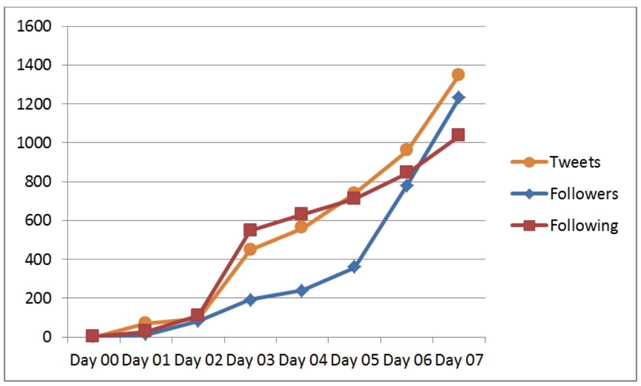

Figure 4: Growth of the agent properties

5. CONCLUSION

The analysis of Web systems is important and essential to design and implement new tools to facilitate the information retrieval by users. In order to know these users, it is necessary to characterize and evaluate their characteristics and extract information about common feelings, group formations and interactions between them.

The use of the SOM neural network to categorize Twitter users proved to be very satisfactory, being possible to characterize and identify some groups by means of user profiles.

Furthermore, it can be seen that the Fuzzy agent represented satisfactorily a human user through a Fuzzy Inference System. The agent carried out a transparent relationship with users, showing typical behaviors of common users. Although users have noticed that the agent was a computerized mechanism, it is possible to conclude that the model showed a reasonable representation of the behavior of a real user, relating to users of different ways and fulfilling its goal of being identified as a member of the group.

As future work we intend to create a refined mechanism for collecting data users to be presented to the SOM neural network, improving and refining the classification process. Moreover, studies should be made taking into account the data users filtered by subject, enabling to identify the most influential users in certain subjects, contributing to the development of new forms of advertising, as well as new content recommendation systems.

6. ACKNOWLEDGMENTS

This research is partially supported by the Brazilian National Institute of Science and Technology for the Web - INWEB (CNPq grant no. 573871/2008-6), CAPES, CNPq, Finep and Fapemig. In particular, the authors would like to thank the INWEB for the assistance in obtaining the real data mass used for the calibration of the neural network.

7. REFERENCES

[1] F. Abel, Q. Gao, G. Houben, and K. Tao. Analyzing temporal dynamics in twitter profiles for personalized recommendations in the social web. *Journal of the American Society for Information Science and Technology*, 2011.

[2] S. G. Arlei Silva, Hérico Valiati and W. M. Jr. From individual behavior to influence networks: A case study on twitter. In *WebMedia*, 2011.

[3] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts. Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM, 2011.

[4] E. Bakshy, B. Karrer, and L. A. Adamic. Social influence and the diffusion of user-created content. In *Proceedings of the 10th ACM conference on Electronic commerce*. ACM, 2009.

[5] C. Bigonha, T. N. C. Cardoso, M. M. Moro, V. A. F. Almeida, and M. A. Gonçalves. Detecting evangelists and detractors on twitter. In *WebMedia*, 2010.

[6] M. Brennan and R. Greenstadt. Coalescing twitter trends: The under-utilization of machine learning in social media. In *SocialCom*, pages 641–646. IEEE, 2011.

[7] P. Brown and J. Feng. Measuring user influence on twitter using modified k-shell decomposition. In *The Social Mobile Web*, volume WS-11-02, 2011.

[8] M. C. Fabricio Benevenuto, Tiago Rodrigues and V. Almeida. Characterizing user behavior in online social networks. *Usenix/ACM SIGCOMM Internet Measurement Conference (IMC), Chicago,USA*, 2009.

[9] Z. Liu, S. He, and W. Xiong. *A Fuzzy Logic Based Emotion Model for Virtual Human.*, pages 284–288. IEEE Computer Society, 2008.

[10] C. Meeyoung, H. Hamed, B. Fabricio, and G. Krishna. Measuring user influence in twitter: The million follower fallacy. 2010.

[11] W. Pedrycz and M. Reformat. *Evolutionary fuzzy modeling*, pages 652–665. Springer-Verlag, 2003.

[12] M. Pennacchiotti and A.-m. Popescu. A machine learning approach to twitter user classification. *Artificial Intelligence*, 46(39):281–288, 2010.

[13] B. Prieto, J. Merelo, A. Prieto, and F. Tricas. *Analyzing a Web-Based Social Network Using Kohonen's SOM*, pages 911–918. 2007.

[14] J. Weng, E.-P. Lim, J. Jiang, and Q. He. Twitterrank: finding topic-sensitive influential twitterers. In *Proceedings of the third ACM international conference on Web search and data mining*. ACM, 2010.

[15] G. P. Zhang. Neural networks for classification: a survey. *IEEE Transactions on Systems Man and Cybernetics Part C Applications and Reviews*, pages 451–462, 2000.