Sentiment classification Method for identification of influential learners in Social Networks Communities

Radhia Toujani[1] and Jalel Akaichi[2]

[1] BESTMOD Department , Higher Institute of management, University of Tunis, Tunisia
toujaniradia@gmail.com
[2] Department of Information Systems, College of Computer Science, King Khaled University Abha, KSA
jalel.akaichi@ku.edu.sa

Abstract. The growth of social networking has gained much interest from the research community in recent years. Social networking technology as an e-learning tool seems promising for education instructors to combine distance education. Several analysis researches of social media were conducted for detection opinion leaders. While most of the existing algorithms proposed for communities determination are destined to commercial use, in this work, we present a new approach for detecting opinion leaders based on analyzing online learning community interactions. In fact, we aim to identify learners behaviors and attitudes in social network sites as productive online tools for learning. To achieve this purpose, we describe a method of performing detecting opinion leaders by using machine learning techniques. We focus on the application of text mining and sentiment analysis. The output of this work prove that education-based social network is very effective and improvement for online communications. Experiments show the efficiency of the introduced method which can be helpful and profitable for education instructors.

Keywords: Social network, learning communities, sentiment classification ,opinion mining, opinion leader, learners behavior, influential learners, machine learning.

1 Introduction

Researchers have become interested to analyse social network site(SNS) and communities detection to explore the educational benefits of using it. Using social network, we construct a virtual communities enable learners to connect and collaborate on international platforms, exceeding geographical boundaries [1], eliminate these boundaries increases communication, collaboration, and engagement [1].

In the shared forum provided by an SNS, learners can construct a shared understanding and engage and collaborate in discussions, while sharing common resources, such as readings, links, and videos [1]. In a distance education course, SNSs provide learners with a venue for fostering and developing a community of practice through technological affordable, such as user profiles, forums, tools, and resources. Given the inherent limitations of course management systems, the use of social network in education settings represents a definitive shift toward social and community-based web applications that cultivate and sustain discipline-specific social networks. For today's technologically savvy students, social network represent improved technologies for creating a heightened sense of community resulting in the acquisition of new student knowledge and collective intelligence [2].

The integration and use of social networking technology as an e-learning tool seems promising for distance education. In [3], the author indicates that social software may be the 'killer app' for distance education, given its ability to enhance

social presence. The creation of social networking around academic topics provides instructors and students great flexibility for teaching and learning[2].
In this paper, we propose a method to detect influential learners in the social networks. At the first step, the extraction and the collection of the raw data are explained. Then, the process of sentiment analysis of online learning communities and the detection of opinion leader are discussed.

The rest of this paper is organized as follows. Section 2 discusses the related works; Section 3 presents the research framework; Section 4 presents experimental results; in Section 5, concludes the paper with some suggestion for future research directions.

2 Related works

Social networks analysis has recently gained more importance thanks to its application in fields such as sociology. Many researchers are interested on identification of opinion leaders. So far, some researchers were studied on opinion leaders where first of them is situated within the field of sociology. Numerous studies such as [5][10][16] have been conducted in order to understand the concept of opinion and characteristics of leaders distinguishing them of their followers.

The studies on opinion leader identification can be divided into two parts. First, the link-based opinion leader identification methods that consider the social interactive structure of the network. Second, the mixture of opinion leader identification methods that combine the social link information with semantic-based information embodied in documents [9][12][15].

As an example of link-based opinion leader selection, the authors in[4] have identified opinion leaders with an excellent edge rank algorithm based on excellent network theory, applied it to rank excellent edges and used the ranking result to identify opinion leaders in opinion excellent network model. Alternatively, researchers in [11][13][17] have proposed an improved combined framework identifying the opinion leader in online learning communities, which ranked opinion leaders based on four distinguishing features: expertise, novelty, influence, and activity. Furthermore, the performances of opinion leaders were further investigated in terms of longevity and centrality.

Also, in[8] [6], authors have found that those opinion leaders are the optimum marketing choice in terms of diffusion speed and maximum cumulative number of adopters, used a social network method and threshold model and conclude that opinion leaders with a high sociality and centrality (users who have high degree of connections to other people) are the best ones for fast diffusion.

In case of social network, many of the trust of customers to the social network communities are based on opinion leaders' recommendations for products and services [7][18]. So, how to identify the opinion leaders effectively is the key to raise sales and brand awareness. In sales and marketing field, some studies focus on developing various indexes such as in-degree, out-degree, betweenness and closeness in social network analysis to identify opinion leaders. These studies are interested in identifying opinion leaders who will forward be marketing messages to other users via their trust and distrust networks.

3 METHODOLOGY

This section presents our methodology. Firstly, the general architecture is presented. Secondly, the extraction and the collection of the raw data are explained. Then, the process of sentiment analysis of online learning communities and the detection of opinion leader are discussed.

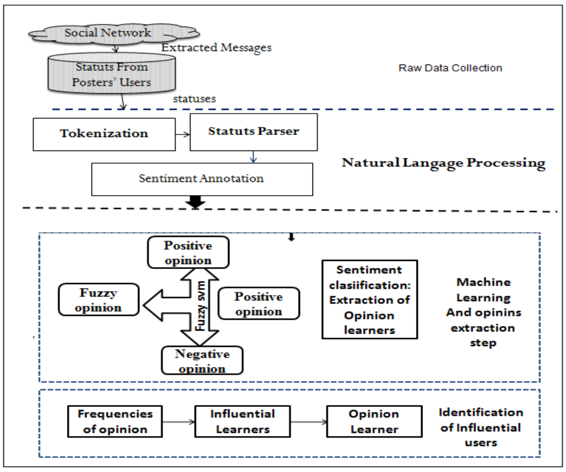

Fig. 1. System Targeted Architecture

3.1 Raw data collection

With the widespread adoption of social media, online learning communities are perceived as a network of knowledge comprised of interconnected individuals with varying roles. Opinion leaders are important in social networks because of their ability to influence the attitudes and behaviours of others. The introduced work is based on the analysis of learning communities in social network namely Facebook. In this study, we chose the type group 'We Read for you'(On a lu pour vous (OLPV))[3]. We used the programming interface available on Facebook site, called Facebook API, providing tools and applications that allowed us to access and get the necessary for information our study. Unlike closed and secret groups, the extraction of database from learning community 'OLPV' group is permitted. This later group has 13,537 members and allows its members to discuss and exchange views on issues related to learning especially reading. Posts the online learning community OLPV are grouped by interest and has many forms namely "status", "photo", "video", "URL" and "event". These posts have multiple messages types:

– An interrogative message where the user requests the opinions of others,
– Quote, poem or proverb message,
– a message where the user expresses his opinion or a message without an opinion.

In fact, these types of posts are shown in the figure 2.

3.2 Machine Learning and Opinions extraction

We introduce machine learning method namely Fuzzy SVM for mining opinions of social network learners. Fuzzy SVM is explored for multi-class scenarios (positive, negative, fuzzy). In fact, we propose a refinement which aims to consider the spring's variables which refer to the membership functions of fuzzy classes. This later is defined as follows:

$$\epsilon_i = \gamma_{w,b} - (Y_i f_{w,b}(x_i))+ \tag{1}$$

where (x)+ represents the positive part of x $\in$ R. This quantity is related to the concept of margin.It measures the location of a pair (X_i, Y_i) in relation to the margin of the classifier f_{wb}. Therefore, three cases are introduced:

[3] https ://www.facebook.com/groups/oalpv.tn/members/

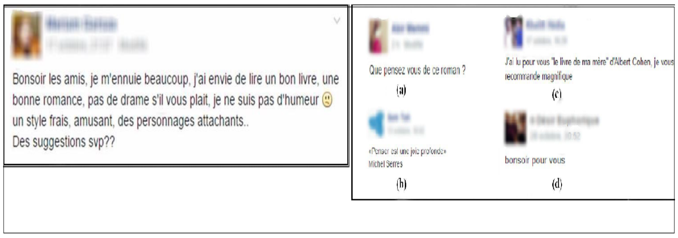

Fig. 2. (a) interrogative message , (b) Quote Message , (c)Message expressing opinion , (d) Message without opinion

- if $\epsilon_i > \gamma_{w,b}$ so the pair (X_i, Y_i) is misclassified by $f_{w,b}$
- if 0< $\epsilon_i < \gamma_{w,b}$ so so the couple is well ranked but close to the hyperplane (the distance from the hyperplane is less than the margin)
- $\epsilon_i = 0$ so the couple is in its majority class (the distance to the hyperplane is greater than the margin).

Indeed, the introduction of these springs variables allow to relax the mathematical formulation of our fuzzy SVM. While the main classes treated on sentiment analysis in social networks are the positive and negative opinion, in our research we focus on fuzzy aspects. To solve this problem, we integrate fuzzy class to labeled input terms. In this context, the training samples are :

$$S = \{(X_i, y_i, s_i), i = 1, ...N \in R^N\} \tag{2}$$

Where each $X_i \in R^N$ denotes a training sample and $y_i \in$ {-1,1} represents its class label. s_i = {1,2,...N} present our fuzzy class with $\sigma \leq s_i \leq 1$ and a constant $\sigma \geq$ 0.Denote a set as:

$$Q = \{X_i | (X_i, y_i, s_i) \in S\} \tag{3}$$

Clearly, it contains two classes. One class contains positive opinion such sample point X_i with y_i=1, denoting this class by C^+, then,

$$C^+ = \{(X_i | (X_i \in S)\} and y_i = 1 \tag{4}$$

The negative class contains such sample point X_i with y_i=-1, denoting this class by C^-, then,

$$C^- = \{(X_i | (X_i \in S)\} and y_i = -1 \tag{5}$$

Indeed,

$$Q = C^+ \cup C^- \tag{6}$$

Thus, the mathematical formulation for our sentiment classification problem and the linear Fuzzy Support Vector Machine classifier can be obtained by solving the following optimization problem:

$$\min 1/2 \|W\|^2 + C \sum_{i=1}^{n} s_i \epsilon_i \tag{7}$$

$$y_i(w^T \Phi(X_i + b)) \geq 1 - \epsilon_i \tag{8}$$

$$\epsilon_i \geq 0, i = \{1, ...N\} \tag{9}$$

where C denotes a constant, the fuzzy class s_i is the fuzzy labeled opinion terms X_i toward one class and the parameter ϵ_i the percentage of opinion class , the term $s_i\epsilon_i$ can be looked as a measure of degree of fuzziness.
Once opinions are detected, a score opinion $ScoreOp$ is assigned . This score is weighted or equal to the number of words contained in the document.

3.3 Detection of influential learners

The detection of online influential learners, we opted for an assessment score corresponding to the frequency of opinion terms contained in each learner message. In addition, to express score of influential learners, we opted a threshold below which the influence of the post is not significant. This threshold is equal to the average of the $ScoreOp$ obtained for all posts.
In other words, any post having a "ScoreOp" higher than or equal to the defined threshold, will be considered as influential leaders. Figure 3 depicts extracted posts from online learning community and presented opinions leaders learners.

Fig. 3. Example of influential learners post

4 SIMULATION RESULTS

Here we present the experimental results of the introduced opinion leaders method on both learning communities $OLPV1$ and $OLPV2$. We use Graph API [4] to extract real time data from these two later communities. In fact, the use of data from the same learner group (OLPV) in two different periods allows us to show the position of opinion leaders over time.
The influential learners belonging to community OLPV1 are shown, in ascending order, in Figure 4

[4] https ://developers.facebook.com/docs/graph-api

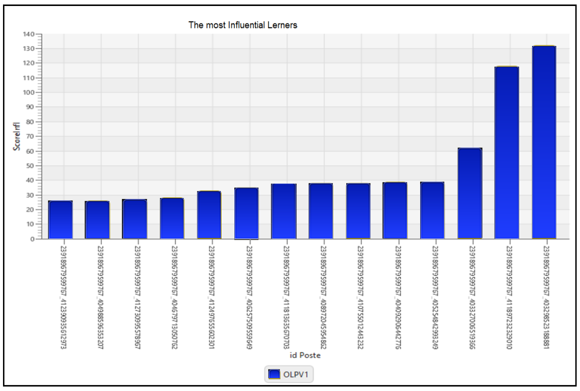

Fig. 4. Diagram of influential learners in OLPV1

Using graph API the ID of the influential learners, we can extract information about the opinion leader's as name and published message. The following figure shows the opinion leader having the highest $ScorInfl$ from the OLPV1 learning community.

Fig. 5. Illustration of name and message of the opinion leader in OLPV1

Figure 6 shows the diagram of influential learner sorted by $ScoreInfl$ of the second learning community OLPV2.

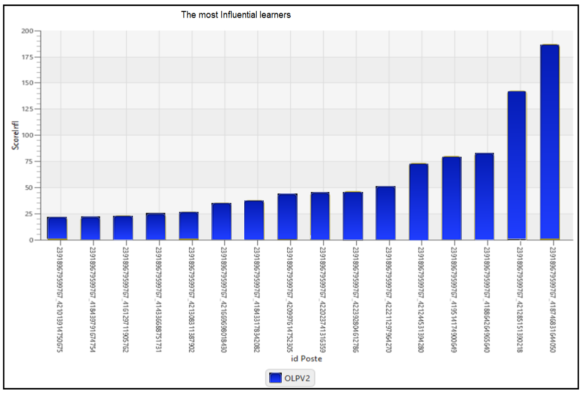

Fig. 6. Diagram of influential learners in OLPV2

Figure 8 depicts information related to influential learner in OLPV2 community.

Fig. 7. Illustration of name and message of the opinion leader in OLPV2

Moreover, we use precision and recall to validate the utility of the introduced method. The precision measures the exactness of a classifier and recall represents the number of correct classifications penalized by the number of missed items. These two measures are calculated as follows:

$$Precision = \frac{I \cap NI}{NI} \tag{10}$$

$$Recall = \frac{NI \cap I}{I} \tag{11}$$

where I denotes the number of outputs influential learners, NI represents the number of outputs non-influential learners. Figure 8 depicts information related to influential learner in OLPV2 community.

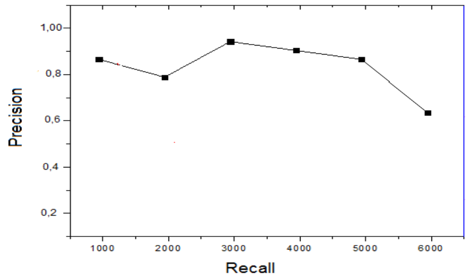

Fig. 8. Illustration of name and message of the opinion leader in OLPV2

According to the curve above, which present the recall and the precision our method gets a higher precision but a lower recall. We can conclude that the performance of the proposed methodology is good.

5 Conclusion

In this work, we have investigated the utility of opinion mining on a novel collection of dataset which is online learning communities. Using the most well-known machine learning algorithm namely SVM, we identify influential learners. First of all, we analyse popularity and reputation of posts to determine which are more important than others referring to the frequency of opinion words in learning posts. Then, we identify influential posts leading to detect opinion leaders learners.
The overall performance of the proposed methodology is satisfactory, however, we would like to further improve our research needs to explore how SNSs can be used most effectively in distance education courses as a technological tool for improved online communications among students in higher distance education courses.

References

1. K. H. McCann., *Virtual communities for educators: An overview of supports and best practices.*, Proceedings from Technology, Colleges, and Community Conference, pages 137142,2009.
2. I. E. Allen and J. Seaman., *Online nation: Five years of growth in online learning.*, MA: Sloan Consortium,,2007.
3. T Anderson., *Distance learningsocial softwares killer ap.*,Adelaide, South Australia: University of South Australia,2005.
4. B. Mullen.et al. , *Shared leadership in teams: an investigation of antecedent conditions and performance.*Acad. Manag. J, 50:12171234,2007.
5. J.B. Carson and al., *Salience, motivation, and artifact as contributions to the relation between participation rate and leadership*,J. Exp. Soc. Psychol, 25:545,559,1989.
6. Y. Cho and al. *Identification of effective opinion leaders in the diffusion of technological innovation: a social network approach.*,Technol. Forecast. Soc, 79:79-106,2012.
7. S. Misra K.K. Chan. *Centrality in social networks conceptual clarification.*,TSoc. Netw, 3:215-239,1979.
8. L.C. Freeman. *Characteristics of the opinion leader: a new dimension.*,J. Advert,19:53-60,2012.
9. J.M. Kleinberg. *Authoritative sources in a hyperlinked environment.*,J. ACM (JACM), 46:604-632,1999.
10. Y. Li and al. *An improved mix framework for opinion leader identification in online learning communities.*KnowledgeBased Syst, 43:34-51,2013.

11. K. H. McCann. *Virtual communities for educators: An overview of supports and best practices.*Proceedings from Technology, Colleges, and Community Conference, pages 137-142,2009.
12. S. Neti. *Social media and its role in marketing.*Int. J. Enterp. Comput. Bus. Syst, 14:1-15,2011.
13. E.C. Nisbet.*The engagement model of opinion leadership: testing validity within a european context*Int. J. Public Opin. Res, 18:3-30, 2006.
14. L. Page and al.*The pagerank citation ranking: Bringing order to the web.*Int. J. Public Opin. Res, 1999.
15. B. Batinic T. Gnambs.*A personality-competence model of opinion leadership.*Psychol. Mark, 29:606-621, 2012.
16. M.P. Venkatraman.*Opinion leaders, adopters, and communicative adopters: a role analysis.*Psychol. Mark, 6:51-68, 1989.
17. G. Weimann and al.*Looking for opinion leaders: traditional vs. modern measures in traditional societies.*Int. J. Public Opin. Res, 19:137-190, 2007.
18. Li.et al. Y.*An improved mix framework for opinion leader identification in online learning communities* KnowledgeBased Syst, 43:43-51, 2007.