Comparative Analysis of Link-based and Content-based Methods for Opinion Mining in Persian Language

Niloofar Allahkaram*
Master of Information Technology Engineering
Science and Research branch of Islamic Azad University
Tehran, Iran
n.allahkaram@yahoo.com

Alireza Yari
Assistant Professor
Iran Telecom Research Center (ITRC)
Tehran, Iran
A_yari@itrc.ac.ir

Received: 2019/04/07 Revised: 2019/06/05 Accepted: 2019/07/16

***Abstract*- Twitter has provided a convenient platform to express feelings and opinions in different areas. Opinion mining in Twitter can be considered as studying the overall sentiment of a tweet. There are two general categories of sentiment analysis methods in the Persian language, linked-base methods and, content-based methods. In this study, we implement a new link-based method for improving opinion classification in the Persian language.**

To compare with the content-based method, we implement a content-based method using Naïve Bayes Method with two different weighting Methods: TF/IDF and Chi-Square. The TF/IDF method has good results in previous Persian language studies. The Chi-Square method has not been used in the Persian language researches, but the accuracy is fairly good in English.

The results show that the improvement in the language-independent methods is remarkable and is in accordance with this research, the precision of the proposed algorithm for positive and negative comments was 98.87% and 97.87%, and the recall value for positive and negative comments was 99.24% and 96.84% respectively. The results also show that because of complexities in Persian syntax and lack of proper natural language processing tools in Persian, content-based algorithms operate poorly compared to English.

Keywords-Opinion Mining; Content-Based; Link-Based;Twitter.

1. INTRODUCTION

In e-commerce, gathering and classifying customer opinions is a major challenge for companies in the context design and marketing, or predicting user reaction to a product or event [8]. Data mining has made it possible to rapidly analyze a high volume of customer opinions. Data mining in social media has become a tool for developing marketing intelligence and predicting developments based on new events or products. Social media data mining can also be called opinion mining [9].

Opinion mining or sentiment analysis is a natural language processing for analysis of attitudes and feelings, or general evaluation of a specific topic, product, or service. Opinion mining is an integral part of text mining that examines textual subjectivity and polarity. In polarity classification, the intention is to discover whether a comment is positive or negative [8].

In recent years, Twitter has become the most popular microblogging social network, and it is used for collection and analysis of comments studied in opinion mining [19]. Many politicians, actors, and celebrities use Twitter to connect, with millions of followers who comment, tweet, like, tag, and share their posts. Twitter Hashtags (#) are a type of tag or label for media data that makes it easy for users to search for content about a specific topic. There are numerous features in social networking applications that are used in opinion mining methods [17].

All previous mining efforts in Persian were language-based. Therefore, this study attempted to implement a language-independent approach based on graph theory for user opinion mining and classification. There are many limitations in the Persian language in terms of natural language processing tools, which are pre-requisites for preprocessing in the lexicon-based method. Hence, this study relies on graph-based methods to bypass existing limitations and improve study accuracy.

The existing methods and literature review are provided in section Two. The proposed method is introduced in section Three. section Four discuss the implementation method and the analysis of the results. Finally, the conclusions were presented in section Five.

2. LITERATURE REVIEW

In the opinion classification literature, there are two general approaches including content-based and link-based methods. Each method involves sub-categories that are shown in Figure 1 [14].

Certain techniques have been used in the Persian language for opinion mining. Sabeti et al. [20]

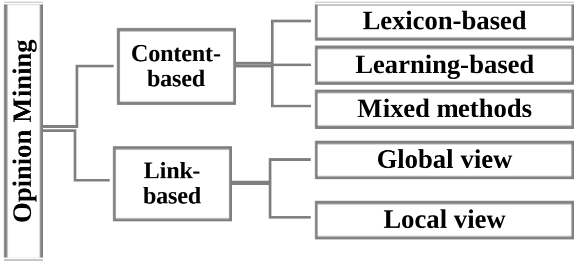

Fig. 1. Methods for opinion classification

introduced a new graph-based method for selecting seeds using dictionaries and expanding it through the k-nearest neighbor (KNN) algorithm and the nearest centroid classifier model. They created an ontology-based sentiment lexicon for Persian called FarsNet using SentiWordnet. The lexicon-based method used in this study relied on translation.

Shirazi et al. [24] proposed a lexicon-based method for determining polarity on Tabnak website comments, which used FarsNet capabilities for mapping Persian synsets to WordNet synsets. By comparing their method with machine learning methods, they concluded that their method is much less costly.

Tavakoli and Rafe [26] presented a method for analyzing the comments left on news websites based on the news content. They attempted to discover the relation of comments or the news to the author's opinion using the syntactic features of the texts, such as nouns and verbs, and analyze the emotional load in sentences.

Mardani and Aghaei [2] used a hybrid method of lexicon and support vector machine (SVM) algorithm for opinion mining in the Persian language. They used SentiWordNet to create a lexicon. Four hypotheses were proposed and tested to improve the study result, with the most effective being the product of repetition polarity and word count.

Bani-Talebi et al. [3] used the naive Bayes machine learning algorithm to analyze comments on social networks, using "bag-of-words" and TF-IDF methods, with the later producing better results.

Shamsfard et al. [23] used data links to classify the polarity of patient experiences. First, a FactNet database was extracted based on the drug polarity, using linked-data and relation extraction techniques. As a result, general semantic patterns were obtained. Then, the knowledge was organized into a hierarchy. Finally, the knowledge was used as a definite entity and general pattern for polarity classification. This method performed better compared to other methods of polarity determination.

Since, there are only a few studies on link-based classification in Persian, two English examples are discussed below. Tan and Tang [25] used link-based algorithm and achieved an accuracy of 76% using local Twitter features, comment similarly in connected users, and the follow feature.

In their 2016 study, Li and Zhu [13] used a new method with global view feature to analyze Twitter data using the retweet feature, achieving 94.35% for "for" user accuracy, 96.74% for "for" user recall,92.79% for "against" user accuracy, and 85.83% for "against" user recall.

Social media provide a new platform for communication that helps determine user opinions. Social network users are connected in a mutual relationship. Hence, opinion mining has become an interconnected classification problem [10].

The problem with rich structured data sets is that the objects are somehow interrelated, which poses a challenge for learning-based methods. In many cases, these data can be described using graphs. Links or edges may display specific patterns that are useful in learning-based methods [6].

Based on the social impact theory, user communication structures can express user comments. Most studies rely on a local view and user comments are classified separately. In such an approach, initial errors may lead to more challenging classifications, thus reducing precision and accuracy. In this study, the proposed method uses the global structure of social interactions to ease the opinion classification problem in a collective manner. Individuals who are discussing a particular subject are modeled as a graph that represents their relational interactions with others. On the basis of the homophily assumption [20], this study models opinion classification as a global consistency maximization (GCM) problem using a link-based model. The collective opinion classifier can find an optimal solution by dividing the graph into two opposing sides, each representing one side of the debate [13].

In addition, this algorithm shows the strength of the training data set. The strength of the link-based approach is in mitigating the common problem of the limited amount of labeled data in big data analytics [13].

This study aimed to improve the accuracy and precision of opinion mining algorithms in the Persian language by introducing a link-based method. These algorithms are not dependent on any language. The proposed algorithm was assessed through two weighting criteria, i.e. TF-IDF [3] and chi-square [20], in the naive Bayes algorithm. These two algorithms are a combination of lexicon-based and learning-based methods.

According to the previous works mentioned in the literature review, most algorithms used in Persian opinion mining are language-dependent and rely on a content-based approach, while there are few studies on link-based algorithms for the Persian language. Therefore, this study aimed to improve opinion mining algorithms for the Persian language using a link-based method. In this research, the algorithm uses a global view in the social media that is split into two parts of the consensus and opposition. This increases the algorithm accuracy with less training data.

3. Implementation Methods

3.1. Language independent algorithm:

According to the Figure 2, In this section we explain the language independent algorithm introduced in this research and the improvements given for the Persian language. As stated above, global consistency maximization is a link-based algorithm that examines user relationships in the form of a graph. Studies on link-based opinion mining are generally based on the Homophily hypothesis. Users who are “connected” by a mutual relationship are more likely to share common opinions [13]. Hence, one node is added to the graph per

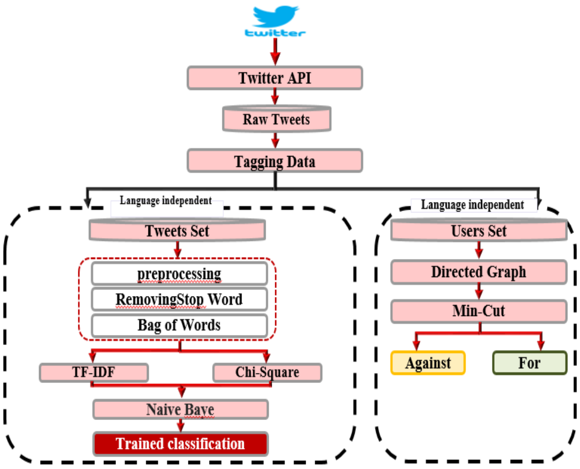

Fig.2. Implementation Methods

user, and each edge represents a relationship between each pair of users. In the context of opinion classification on a social media platform like Twitter, researchers have investigated users' mutual relationships in a variety of forms, including "follow", "mention," or "retweet." [21]. These relationships can be indicative of shared views between users. However, several studies suggested that the "follow" links had a little positive impact on the classification accuracy [13]. The proposed model explores a set of users and their relationship in terms of "retweet".

For classification training, the users in dataset are divided into two categories. Users who have clear comments that are considered as training model seeds and those whose views are unclear are considered as test data. In this study, a number of influencers from both sides of the election debate plus a number of users with over 1,000 followers and a relatively high number of retweets were used as "seed". Users who retweeted the original tweets and had tweeted 2-4 times according to [13] were used as test data. When drawing the graph, the user stance is ignored, until later when it is determined by the min-cut. The following section explains how the min-cut works and how the graph is drawn in the proposed method Equations.

- Minimum Cut

Min-cut is an optimization method. In graph theory, a cut is removing a number of connected graph edges to divide the graph into two disjoint subsets. If removing edges had a cost, the minimum cut involves finding a subset of edges that minimizes the cost of removing these edges. The cost of a cut in unweighted graphs is equal to the number of edges crossing the cut, and in weighted graphs, the sum of the weights of edges crossing the cut. A s-t cut on the graph G with two terminals will divide the nodes into two S and T subsets, such that the source of s is S and the source of t in T. The minimum cut in graph G can be the smallest s-t cut in G.

- Graph

In this way, the mutual relationship between users and the opinion classification are presented in a global view. First, a directed graph of users involved in an online discussion is drawn. This approach finds the optimal solution by dividing the graph into two subsets including the sides of the discussion.

Then, the social network is represented with a directed graph G(V, E). In such a graph $v{:}\{v_1\, v_2 \cdots v_n\}$ is a set of n nodes, where each node represents a user. $E{:}\{e_{ij}\}$is a set of edges, where each edge represents a relationship between the users i and j. When users discuss a certain topic, user i may hold an opinion state x_i. Here, it is assumed that only two possible opinion states exist. Hence $x_i = 1$ if user *i* is a supporter "for", and $x_i = -1$ if user *i* is an opponent ("against"). If the user stance is not clear, it is expressed as $x_i = 0$, and eventually, a value of 1 or -1 is assigned to it. The data was labeled manually, where tweets were divided into three categories: positive (agreement), negative (disagreement), and neutral (e.g. declarative tweets). Users classified under the neutral category were removed from the dataset. Users classified under positive and negative categories with over 1,000 followers and acceptable retweet rates were used as seed (training data) expressed by v_v. Other users from these two categories with 2-4 tweets were considered as users with uncertain views (test data).

Each edge represents retweet relationship between two users; hence, the weight of the edge between two users, w_{ij}, is the number of retweet action between them (user *i* retweeting user *j* and vice versa). According to (1), for each pair of users *i* and *j* with opinion states $x_i$and x_j, $w_{ij}x_ix_j$ is defined as a consistency score of the edge connecting *i* and *j* [14]:

$$\text{Maximize } E = \textstyle\sum_i \sum_j w_{ij}x_ix_j \qquad (1)$$
$$s \cdot t \cdot \qquad x_ix_j = \{1 - 1\ 0\}$$

Here, the directed graph G has two new nodes in addition to the user nodes: a source node s (from) and the target node t (to). Therefore, it is necessary to have rules for drawing the graph. For each node i in G with $x_i = 1$, a directed edge is drawn from node s to node i with $w_{si} = \infty$.

For each node i in G with $x_i = -1$, a directed edge is drawn from node i to node t with $w_{it} = \infty$. For users i and j connected such that $x_i = 1$ and $x_j = 0$, a directed edge is drawn from node i to node j with the weight w_{ij}.

For users connected such that $x_j = 0$ and $x_i = -1$, a directed edge is drawn from i to j with the weight w_{ij}. For users i and j connected such that $x_j = 0$ and $x_j = 0$, a directed edge is drawn from *i* to *j* and another from *j* to *i* with $w_{ij} = w_{ji}$. In this graph, edges that are not on either node ends in state 0 are ignored.

With this directed graph G, classifying user opinions is converted to a min-cut problem. The s-t cut in graph G is a partition of the nodes into two subsets S and T, where S [13]. contains the source node s and T contains the target node t. An edge (u, v) crosses the cut if u lies in S and v lies in T. The weight of the cut is the sum of the weights of the edges crossing the cut. The objective is to find the s–t cut C with the smallest weight in G. For each node i in S, $x_i = 1$, and for each node i in T, $x_i = -1$. Since the weights of the edges incident on s and t (edges of types 1 and 2) are set to be infinity, they cannot be selected to participate in C.

Therefore, only edges incident on nodes with state 0 are selected to participate in C. Edges in C for users i and j, with $x_i = 1$ and $x_j = 0$, or vice versa, correspond to users with opposite opinions who have retweeted one another. Users with $x_j = 0$ and $x_j = 0$ have one C member edge at most; since, according to s-t cut, edges with one corresponding edge in the opposite direction from a node in the set T to a node in S, does not belong to C. Finally, the cut C with the smallest weight yields the optimal state that minimizes the total weight of inconsistent edges in G.

In the example in Figure 3, suppose the weights of all edges not incident on s or t are 1 and the min-cut in G is the edge connecting the node in state 0 to the one in state −1. Thus, the algorithm should assign a state of 1 to both nodes in state 0.

3.2. Content-based algorithms

The main objective of this research is to improve the methods of theoretical analysis in the Persian language. Therefore, to find a proper assessment benchmark, it was necessary to implement two content-based algorithms, where the naive Bayes learning algorithm was used for both these methods. These two algorithms weight words differently as explained below.

- Naive Bayes Algorithm

This algorithm is a probabilistic classification that learns the pattern of classified documents and compares the content with a list of words to classify the documents in appropriate classes. If *t* denotes a tweet, and C * is the class to which *t* should belong, then, the formula is [1]:

$$\mathbf{C^* = argmac_c P_{NB}(c|t)} \tag{3}$$

$$\mathbf{P_{NB}(c|t) = \frac{(P(c)) \sum_1^m P(f|c)^{n_{i(t)}}}{P(t)}} \tag{4}$$

In the above equation, *f* is a feature and *f (i)* is the number of features that are displayed with $n_{i(t)}$, and *t* represents a tweet. Here m has no features. The $P(C)$and $P(f|c)$ parameters are calculated through maximum likelihood estimation, and it is used for smoothing invisible features.

- Chi-square method

Chi-square [20] is a weighting method used in this study, which has not been implemented in the Persian language before. In this method, the number of words repetitions is calculated; the total repetitions and the per class repetitions. The total repeat distribution is calculated for all words, and the conditional repeat distribution is calculated for tagged classes. Then, with the help of these numbers and the chi-square function, the words are ranked and words whose score is greater than 3 are selected according to [20]. These words are then placed in a set and tested using the feature function. Then, tweets are classified according to the presence of these words. The data is trained using the naive Bayes algorithm.

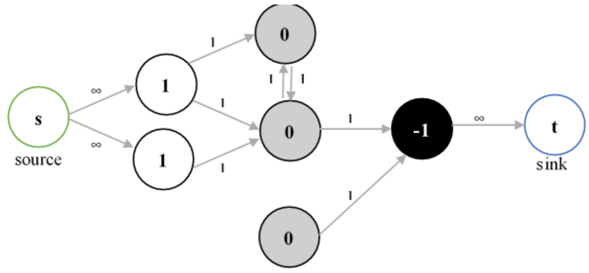

Fig.3. Example directed graph [14]

According to (5) and (4), We implemented the chi-square method.

$$\text{chi}^2 = \sum_{t=1}^{k} \frac{(O_t - E_t)^2}{E_t} \tag{5}$$

- TF-IDF method

Another algorithm used for comparison is the TF-IDF weighing method [3]. It is the product of term frequency and inverse document frequency, indicating the relative importance of a term in the document and the corpus. First, the ratio of the normalized value of term repetition in the document is calculated per the total number of words in the document. In this study, TF-IDF was calculated at word level. After weighing the words, the obtained weight and the word itself are taken as a feature of the text and then classified using a naive Bayes algorithm. According to (6) and (4), We implemented the chi-square method.

$$\text{tf} - \text{idf}(t_k, d_j) = \text{tf}(t_k, d_j) \times \log \frac{|N|}{N(t_k)} \tag{6}$$

4. EVALUATION

Only a set of user IDs and those who retweeted them is required for the implementation of the GCM algorithm. The data was obtained using a program in Visual Studio and SQL in the JSON format. The NetworkX library is used to draw the directed graph with the min-cut method. SkitLearn library is also used in the Python language for training and classifying the naive Bayes classification.

4.1. Simulation Environment

This study was conducted using Ubuntu 14.06.3 OS with Intel® Core ™ i7-8550u processor and 4GB of RAM installed on the VMware virtual machine. A Python-built application programming interface was used to collect tweets. A specifically developed application, written in C#, was used to tag the tweets. The Python language was used to clean up and implement the classifications. ATOM 1.28.1 editor was used for Python codes

4-2. Datasets

The training dataset contained 110 users, including 87 "for" users and 39 "against" users, and a total of 4954 tweets that were collected from Twitter and then tagged. Meanwhile, of 970 users with 2-4 tweets who had retweeted the seed users, 500 were randomly selected, which consisted of 266 "for" users and 95 "against" users; the rest were not fit to be included. As a result, 361 users made up the test data, with a total of 769 "for" tweets and 277 "against" tweets.

5. Assessment

In previous studies, the accuracy and recall and F-measure criteria were used to assess the proposed systems. In this section, the results of the study are analyzed. Considering that the content-based and the third algorithms are language-independent, the same training and test sets were used for all three algorithms in order to provide a reliable scale for comparison of the three algorithms. According to Table 1, the accuracy and recall and F-measure for "for" and "against" data using the three algorithms can be examined separately in Diagram 1, 2 and 3.

This section compares the results obtained from the implementation of the GCM algorithm and the results from [14]. Table 2 shows the accuracy and recall and F-measure results for "for" and "against" data.

The results obtained from both systems are very similar, however, the proposed system has performed slightly better for both accuracy and recall and F-measure results. This was due to shortage of seeds for random selection.

Another of our comparisons is the result of the implementation of two Methods of Chi-square in Persian and English.

According to the diagram 4, this algorithm has not worked well in Persian, due to the lack of Persian language preprocessing tools and the complexity of Persian language with respect to English.

6. Conclusion

Sentiment analysis, opinion mining, and entity analysis are interrelated fields of research that use various techniques such as natural language processing, data retrieval, and structured and non-structured data mining. This study employed a link-based opinion mining algorithm to improve the classification of Persian comments. This algorithm was compared with two mixed learning-based and learning-based algorithms in the Persian language. Twitter data relating to Iran's 2017 presidential election was used to test these methods. The contribution of this study is introducing a link-based algorithm which is language-independent and can improve the classification of comments due to the complexity of the Persian language. The results showed a significant improvement and the advantage of the proposed link-based method over similar studies in the Persian language, which used lexicon-based and learning-based methods.

TABLE 1. TRAINING AND TEST DATA USERS

		For	Against
Training Dataset	Number of Users	87	33
	Number of Tweets	3773	1181
Test Dataset	Number of Users	266	95
	Number of Tweets	769	277

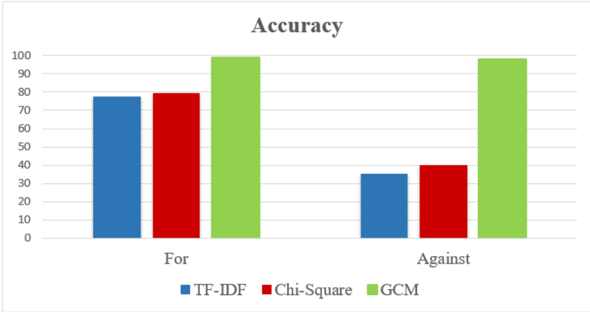

Diagram 1. Accuracy for three algorithms

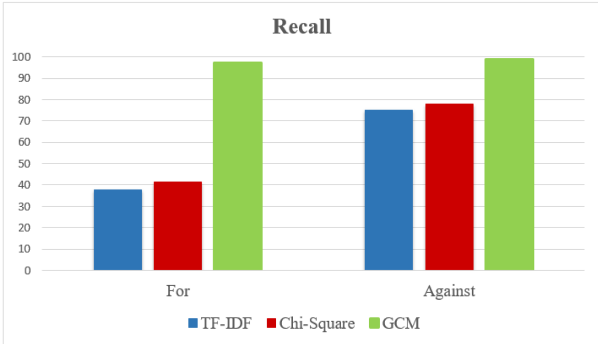

Diagram 2. Recall for three algorithms

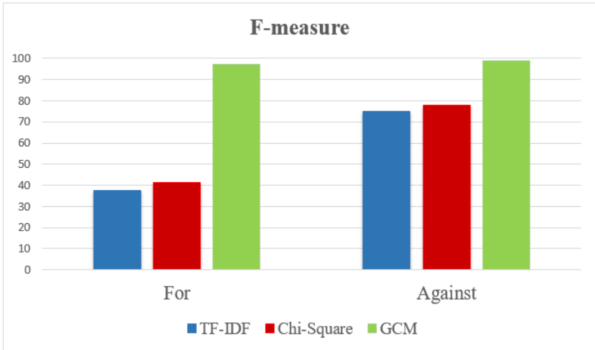

Diagram 3. F-Measure for three algorithms

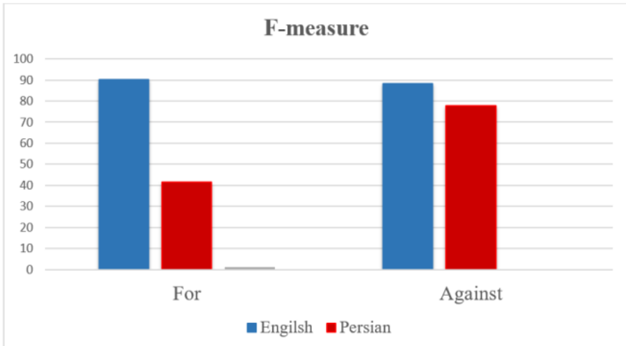

Diagram 4. F-Measure for Chi-Square algorithms.

TABLE 2. COMPARISON OF GCM IN PERSIAN AND ENGLISH

	For			Against		
	Accuracy	**Recall**	**F-measure**	**Accuracy**	**Recall**	**F-measure**
Proposed System	98.49	96.62	97.32	96.84	98.10	99.05
GCM	94.35	96.74	95.53	92.79	85.83	89.18

Also, if we compare the result of chi-square algorithm in Persian and English, we find that content-based algorithms do not work very well in Persian due to the complexity of the language and the lack of pre-processing tools. Although this algorithm is in good agreement with the algorithm TF-IDF in Persian, according to Diagrams 1 and 2 and 3. Indeed, new weighing methods may improve content-based algorithms, but still is as effective as link-based algorithms in Persian.

In the future, a combination of link-based and Content-based algorithms can be used to improve language-dependent algorithms in the Persian.

REFERENCES

[1] Adinarayana, S. & Ilavarasan, E. (2017). December. An efficient approach for opinion mining from skewed twitter corpus using over sampled imbalance data learning. *In 2017 International Conference on Intelligent Communication and Computational Techniques* (ICCT), (pp. 42-47). IEEE.

[2] Alimardani, s., & Aghaei, A. (2015). Opinion mining in Persian language using supervised algorithms. Journal of Information Systems and Telecommunication. 3. 135-141. 10.7508/jist.2015.03.001

[3] Banitalebi, A., dr. Zamani F (2016). Social Networking Opinion Mining Using Machine Learning Algorithms. International conference on Computer Engineering & IT in Persian.

[4] D'Avanzo, E. and Pilato, G. (2015). Mining social network users opinions to aid buyers' shopping decisions. *Computers in Human Behavior*, 51, 1294-1284.

[5] Eisenstein, J. (2016). Unsupervised Learning for Lexicon-Based Classification, *In Thirty-First AAAI Conference on Artificial Intelligence*.

[6] FasihFar, Z., RamezanPour, S., Izzy, H., & Rokhsati, H. (2017). Review on the application of semantic relationships and a combination method in Opinion Mining and emotion Analysis in the Twitter social network. *The 2nd International Conference on Knowledge Based Research in Computer Engineering and Information Technology*, Tehran, University of Journalism in Persian.

[7] Golpar-Rabooki, E., Zarghamifar, S., & Rezaeenour, J. (2015). Feature extraction in opinion mining through Persian reviews, *Journal of AI and Data Mining*, 3(2), 169-179.

[8] Hosseini, Sh. (2016). Review article on the Mechanisms Provided in Opinion Mining. *Second National Conference on Computer Engineering Research*, Hamedan, Ekbatan Research Group in Persian.

[9] Imani, s., Ali Ashraf Payaman, N., & Khodaee, A. (2016). Investigating Methods for Using Sense Analysis Algorithms in Social Networks. *Second International Conference on New Findings of Science and Technology*, Qom, Islamic Studies and Research Center, Soroush Hekmat Mortazavi in Persian.

[10] Kahriz, j., Abdi kharabe, R., & Ghahramani. S. (2017). Reviewing the Opinion Mining on Social Networking. *5th International Conference on Electrical and Computer Engineering with Emphasis on Indigenous Knowledge*, Tehran, Holy Ardabil Institute of Higher Education in Persian.

[11] Kahzad, M., Zamani Brojeni, F. (2017). Review of Emotion Analysis in Persian Language . *Sixth National Conference on New Ideas in Electrical Engineering* in Persian.

[12] Kouloumpis, E., Wilson, T., & Moore, J. D.(2011). Twitter sentiment analysis: The good the bad and the omg!. *Icwsm*, 11(538-541), 164.

[13] Li, J., Li, X., & BinZHu, Z. (2016). User opinion classification in social media: A global consistency maximization approach, *Information & Management*, 53(8), 987–996

[14] Li, Y. M., & Li, T. Y. (2013). Deriving market intelligence from microblogs. *Decision Support Systems*, 55(1), 206-217.

[15] Lu, Q., Getoor, L. (2003). Link-based classification using labeled and unlabeled data. In *ICML 2003 workshop on The Continuum_from Labeled to Unlabeled Data in Machine Learning and Data Mining*, V.3, PP. 496-503.

[16] Mumtaz, D., & Ahuja, B. (2016). A lexical approach for opinion mining in twitter. *International Journal of Education and Management Engineering* (IJEME), 6(4), 20.

[17] Peykari, N., Yaghobi, A., & Taheri, H. (2015). Emotions analysis in the social network of twitter with the technique of text mining. *1st International Conference on Web Research*, Tehran, University of Science and Culture in Persian.

[18] Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-Based Systems*, 89, 14-46

[19] Sabeti, B., Hosseini, P., Ghassem-Sani, G., & Mirroshandel, A. (2016). LexiPers An Ontology Based Sentiment Lexicon for Persian, GCAI. *2nd Global Conference on Artificial Intelligence*, 41, 329-339.

[20] Sayeedunnissa., S. F., Hussain, A. R., & Hameed, M. A. (2013). Supervised opinion mining of social network data using a bag-of-words approach on the cloud. In *Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA (2012).* (PP.299-309). Springer, India.

[21] Shahidi, N., Ahmadi Abkenari, F., Mirroshandel, A. (2017). Recognizing the template in the users' review process using the label words and polarity of opinion. *Third International Conference on Pattern Recognition and Image Analysis*, Shahrekord, Shahrekord University, Iran Visual Image Processing and Image Processing Association in Persian.

[22] Shams, M., Shakery, A., & Faili, H. (2012, May). A non-parametric LDA-based induction method for sentiment analysis. *16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP 2012)*. (pp. 216-221), IEEE.

[23] Shamsfard, M., Noferesti, S. (2015). Using linked data for polarity classification of patients experiences. *Journal of Biomedical Informatics*, 57, 6–19.

[24] Shirazi, H., Aarafi, M., SadidPour, S. (2015). Determine polarity by calculating the mean of polarization values of sequence sets. At the *2nd International Congress of Iranian New Technologies to achieve development*. Center for Paydar Development, Sustainable Institute. Higher education in Arvand in Persian.

[25] Tan, C., Lee, L., Tang, J., Jiang, L., Zhou, M., & Li, P. (2011, August). User-level sentiment analysis incorporating social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining.* (pp. 1397-1405). ACM.

[26] Tavakoli, A., Rafe, A. (2016). Presentation of a method for the emotions analysis in the context of comments. *The first national conference on interdisciplinary research in computer engineering, electrical engineering, mechanics and mechatronics*. Qazvin, Beyon Zahra Technical Engineering Center, Qazvin Province Science and Technology Park in Persian.

[27] Lo, Y. W., & Potdar, V. (2009, June). A review of opinion mining and sentiment classification framework in social networks. In *2009 3rd IEEE international conference on digital ecosystems and technologies*, (pp. 396-401). ieee.

Niloofar Allahkaram received her B.S. degree in Information TechnologyEngineering from Kordestan university, Iran, in 2013 and her M.Sc. degree in E-commerce engineering from University of Science & Technology, Tehran, in 2019 .Her research interests include Persian sentiment analysis in social networks.

Alireza Yari received his B.Sc. degree in control system engineering in 1993 from the University of Tehran, Iran, and M.Sc. and a Ph.D. degree in System engineering in 2000 from Kitami institute of technology, Japan. He is currently doing research in the Information Technology research faculty of ICT research Institute of Iran. His research interests include web information retrieval and language-specific search engine, e.g. Persian search engine. He is also working on Persian sentiment analysis in social networks.