Unsupervised aspect-based Sentiment Analysis in the Persian language: Extracting and clustering aspects

Conference Paper · October 2020

DOI: 10.1109/ICCKE50421.2020.9303651

Kourosh Hashemi
Ferdowsi University Of Mashhad

Unsupervised aspect-based Sentiment Analysis in the Persian language: Extracting and clustering aspects

Reza Akhoundzade
Sharif University of Technology
Tehran, Iran
Reza.akhoondzade@ce.sharif.ir

Kourosh Hashemi Devin
Ferdowsi University of Mashhad
Mashhad, Iran
Kurosh.hashemi@mail.um.ac.ir

Abstract— Sentiment analysis is the field of natural language processing to analyze user feedback, preferences, and evaluations from the text. Different organizations in most social scopes use this context as an appropriate tool to find their strengths and weaknesses. In this field of research, the objective is to determine the positive or negative orientations of users towards the features of a product or commodity. Solving this problem consists of two main steps: aspect extraction and identifying the positive or negative tendencies of users towards those aspects. Two of the most critical issues of sentiment analysis in the Persian language are the lack of comprehensive labeled data and the significant difference between colloquial and formal sentences in Persian. One of the most important methods to confront the first problem is to apply unsupervised methods. In this research, a system for aspect-based sentiment analysis in the Persian language is proposed using unsupervised methods. The Sentiment words extraction step is done in another article [1]. The aspect extraction and clustering steps are done using topic modeling methods and neural networks and taking benefit of rule-based methods. This system has been evaluated using precision and recall criteria. The F1 criterion for extracting aspect words in the proposed system was 0.766.

Keywords— sentiment analysis; Aspect-based; learning methods; neural networks; Aspect extraction; Aspect clustering; unsupervised methods; precision; recall

I. INTRODUCTION

In recent years, along with the increase in the volume of textual data at the web level, some studies have been conducted in the field of automatic analysis and evaluation of this textual data. A large amount of this textual data is the user's feedback. Therefore, there is a need for a system that can automatically analyze and evaluate these opinions. Sentiment analysis is a field of natural language processing that analyzes users' feedback, preferences, evaluations, and feelings [2]. This field is one of the most active branches of natural language processing, the importance of which has increased with the ascending growth of social networks, and their textual contents.

In research, comments were used to rate goods [3]. Another study examined the link between Twitter sentences and public opinion polls. Other studies have used Twitter data and movie reviews to guess the sales of movies [4-5]. There are several other applied studies in this field that demonstrate the extension of sentiment analysis. [6]

Sentiment analysis has many challenges. The challenges of sentiment analysis are divided into two general categories: (1) natural language processing challenges, and (2) Sentiment Lexicon challenges. One of these challenges is to be able to deal with the written text in a conversational method.

One of the most important steps in the sentiment analysis problem is to recognize sentiment words. These are the words that are commonly used to give a positive or negative opinion. The list that includes these words, phrases, and terms is called the Sentiment Lexicon. To best of our knowledge, there is no complete Lexicon available for Farsi. In another article, we have extracted and produced a Sentiment Lexicon [1]. In this paper, the same Sentiment Lexicon is used to perform the next step of sentiment analysis. Of course, in the case of sentiment analysis, this dictionary alone is not enough. In this paper, our center of attention is on the next step, i.e., the extraction and clustering aspects, and the aspect-sentiment classification.

Section 2 examines previous works. Section 3 discusses the proposed method. Chapter 4 considers the obtained results.

II. BACKGROUND

This paper only investigates previous researches on the steps of extracting and clustering aspects and the aspect-sentiment classification. This step includes aspect extraction, or opinion target, and determining the interrelation of the aspect and the sentiment word. This problem can be considered as an issue of information retrieval. The main rule of sentiment analysis is that each sentiment word such as "good", "bad", etc. must have an aspect target. Therefore, it is important to be able to extract aspects or features automatically.

A. There are four main methods to extract aspects:

1. Extraction based on frequent names and phrases.
2. Extraction based on sentiment and aspect relations.
3. Extraction using supervised methods.
4. Extraction using topic modeling and neural networks.

Hitherto, there are many methods to extract aspects, which are discussed below. One of the main challenges in aspect extraction is the lack of labeled data in all areas and scopes of application. The problem of the lack of labeled data in Persian is much more than in English. Therefore, it was decided to apply unsupervised methods in this research.

Unsupervised methods of dealing with this problem are topic modeling methods and neural networks. Therefore, in this article, only extraction methods based on sentiment and aspect

978-1-7281-8566-8/20/$31.00 ©2020 IEEE

relations and unsupervised methods will be studied and discussed. It is also possible to offer a combination of these methods for extracting and clustering aspects. The final proposed method of this research is a combination of methods 1, 2, and 4, as mentioned above.

B. Extraction based on sentiment and aspect relations

The basic rule of this problem is that every comment must be relevant to one aspect. Therefore, there must be a connection between the aspect and comment. Sometimes there may be a few words in a sentence, but there is no repetition aspect in the sentence. In this case, the nearest noun or phrase can be considered as an aspect [7]. The dependency parse tree is also used to recognize the interrelation of aspect and comment [8-9]. The idea of the dependency tree has been used in other research work as well [10].

The parse tree cannot be used in Persian because there is no proper tool for extracting the dependency tree of Persian colloquial sentences. Hazm[1] tools also cannot have the ability to analyze colloquial sentences.

C. Extraction using topic modeling and deep learning

The subject that has been seen in several research studies is that, due to the inability of completely unsupervised topic modeling methods for aspect extraction, in most of the presented researches and methods, a generalization of topic modeling methods has been used to solve the problem. All of the methods presented in recent years are semi-supervised, which, as discussed, have their specific issues. But the policy of this research is that the extraction of aspects should be done in a completely unsupervised method.

Many recent research studies have used deep learning and neural networks, especially the Word2Vec [11] model, to extract and clustering aspects.

Pavlopoulos et al. used a text-based database, including user feedback, to develop a Word2Vec model [12]. English Wikipedia has also been used to teach another model, which is used to remove stop words. Then, using the original Word2Vec model vectors and combining it with the method provided by one of the previous works [13], the aspects have been extracted.

In other research, a deep learning system was introduced to analyze users' sentiments on Twitter. The main difference is the use of another model for the initialization of neural network weight parameters. Other researches have been conducted using neural networks on the Mandarin and the Sina Weibo website, known as Chinese Twitter [14]. Another study also used clustering methods to cluster the views of Twitter and Sina Weibo. [15]

Several other studies use the Word2Vec vector in their semi-supervised method. In a 2016 study, the problem of aspect-based sentiment analysis was performed on laptop and restaurant scopes [16]. In this research, some aspects are considered as the input of the system, which makes the proposed system of this research semi-supervised. The extracted features are given as input to a supervised clustering method. The method presented by this paper on the restaurant and laptop scopes is 77.81% and 76.81%, respectively.

As observed, most of these studies, using the neural networks provided by the article itself, or the previously described Word2Vec output vectors have solved important sub-problems of sentiment analysis. It is noteworthy that, with the use of clustering methods, attempts have been made to categorize user comments. The cosine similarity criterion is used in most of these cases.

A brief description of some of the researches conducted in Persian is provided as follows.

In research conducted by Pedram Hosseini et al.[17], they provided a Persian dataset for the problem of sentiment analysis, called the SentiPers. The dataset contains 26,000 user comments on digital goods. In this study, these datasets have been used to evaluate the final proposed system.

In research, a new method is proposed by Khiabani el al. [18] for scores aggregation that employs both the most and the second probable classes to predict the final score. The proposed approach considers every review as a set of sentences each of which has its own sentiment orientation and score and computes the probability of belonging of every sentence to different classes on a five-star scale using a pure lexicon-based system. Their proposed method is applied to review datasets of TripAdvisor and CitySearch which have been used in previous studies which makes the precision of the proposed method for both datasets get 23% and 27% higher, respectively.

In research, the sentiment classification process for the Persian language is performed by deep learning method. Dastgheib et al [19]. proposed a hybrid method by a combination of structural correspondence learning (SCL) and convolutional neural network (CNN). The SCL method selects the most effective pivot features so the adaptation from one domain to similar ones cannot drop the efficiency drastically.

In another work, Dashtipour et al. [20] propose a novel hybrid framework for concept-level sentiment analysis in the Persian language, that integrates linguistic rules and deep learning to optimize polarity detection. When a pattern is triggered, the framework allows sentiments to flow from words to concepts based on symbolic dependency relations. When no pattern is triggered, the framework switches to its sub-symbolic counterpart and leverages deep neural networks (DNN) to perform the classification.

In another research, Shams et al. [21] proposed an unsupervised paradigm for aspect-based sentiment analysis, which is not only simple to use in different languages, but also holistically performs the subtasks for aspect-based sentiment analysis. Their methodology relies on three coarse-grained phases which are partitioned to manifold fine-grained operations. To determine the polarity of any aspect in the final phase, the document is firstly broken down to its constituting aspects and the probability of each aspect/polarity based on the document is calculated.

In this article, we have tried to make all the steps of solving the problem of sentiment analysis at the aspect level as

[1] Set of natural processing tools for Persian

unsupervised as possible. In the next chapter, the proposed method of this research will be described.

III. The Proposed Method

In order to solve this problem with unsupervised methods, a large amount of textual data must first be collected on the subject. As we all know, in unsupervised methods, we do not need labeled data, but to solve the problem using machine learning methods, we need a large volume of raw data that needs to be pre-processed. Also, using these data, a Word2Vec model has been trained using neural networks. These steps are fully covered in another article, and this article deals with the extraction and clustering aspects step.

The proposed solution for the extraction of aspects is to apply a combination of rule-based methods and neural networks using the Word2Vec model [11-12]. In this study, firstly, the candidate aspects are extracted using rule-based methods. Then, the most frequent candidate aspects are selected, and the final aspects are extracted using clustering methods and the criteria of cosine similarity. Figure 3.1 shows the schema of our proposed method. In the following, we discuss the details of the proposed method.

After extracting the sentiment words, the aspects or the features of the desired product should be searched. If we assume that the main features of a commodity have been given, the rest of the features of the commodity also can be extracted by cosine similarity without any challenges. For example, if we have the main features of a mobile phone, such as "battery" (batri)[2], "sound quality" (keyfiat seda), "antenna" (antendehi), "price" (gheymat) and "camera" (dorbin), we can very simply apply the cosine similarity criteria to other features of this product. Other features include words such as "battery" (battri), "speaker" (bolandgoo), "battery" (battrish), "speaker" (spiker), "price" (gheymatesh), "sound" (sedash) and other words that are used in colloquial language. But in this study, the objective is to automatically extract aspects without the need for prior knowledge and key features.

In this research, firstly, the candidate aspects are extracted using an adjective and descriptive rules based on the grammar. Obviously, this method does not recognize all aspects of the product. It is also possible to extract the wrong aspect. However, it is noteworthy that the aspects obtained at this step are just a starting point for extracting and clustering other aspects using clustering methods.

In general, in Persian, candidate aspects are nouns or noun phrases, which come before a positive or negative adjective. They may also have been preceded by two consecutive adverbs and adjectives, such as "very excellent." It can be stipulated with a high probability that an aspect word in its nature, comes in one of the two forms mentioned in the text. The solution used in this research is that, in order to extract candidate aspects, using the tags of the part of speech, we only look for phrases such as "good phone" (gooshie khoobie), "great camera" (dorbin awli) and "vary great phone" (gooshie kheyli awlie).

The important point is that from the aspects extracted in this method, it is possible to eliminate cases that have a low frequency. The rest of these aspects, which are candidate aspects, have been selected as aspects derived from the rule-based method, but are not yet final aspects. Aspects that have been repeated less than 50 times have been removed at this step.

It is noteworthy that the candidate aspects obtained in this section are the inputs of the next step so that in the next step, all the desired product aspects are extracted and clustered, using the clustering method and output vectors of the Word2Vec model.

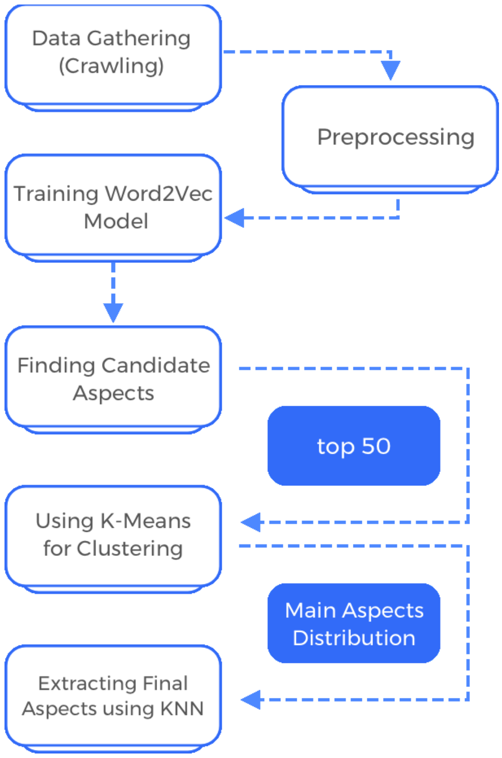

Fig. 3.1 Schema of our proposed method

A. *Extracting and clustering final aspects using clustering methods*

This is the most important step. The purpose of this step is to attain the main aspects (header aspects) so that by using them, the rest of the aspects can be extracted and clustered. At this step, the most frequent candidates are first selected from among the aspects obtained in the previous step. By the most frequent, we do not mean the common words used in the text. It means the nominal expressions with a large number of repetitions in adjective and descriptive forms.

In the proposed method of this article, 50 of the most frequent candidate aspects are selected where they were extracted in the previous step. These 50 aspects and their corresponding Word2Vec vectors are then given as K-Means inputs. The main advantage of K-Means output in this section is that it can demonstrate to us the distribution of the main aspects in the vector space with proper accuracy. It can be mentioned that the output clusters of K-Means each include one of the main aspects such as "phone", "battery", "camera", "sound quality", "price" and so on. Also, the mean of each of these clusters can

[2] The words in parenthesis are the transliteration of Persian words having the same meaning as those before the opening parenthesis.

be considered as a representative vector for one of the main aspects or the header aspects.

The K-Means output cluster intermediaries represent the main or leading aspects that have been explored in this study. Using these intermediaries as the center of clusters, all aspects close to these centers can be extracted and clustered. The aspects close to the center of the clusters are the same as the final aspects.

Finally, the final aspects include 50 of the most frequent aspects and the rest of the candidate aspects that are close to these frequent aspects, or in other words, the K-Means output clusters. In this section, the cosine similarity criterion is used for the closeness of words' meaning. Each candidate aspect, where its similarity to each of the cluster centers is higher than the set threshold value, is extracted as a final aspect. It also enters the desired cluster as a co-cluster, and clustering is performed simultaneously.

The purpose of this step was to select only those aspects that would be close to one of the final aspects. For example, the words "battery" (batri), "price" (gheymatesh) and "appearance"(zaher) have been among the top 50 most frequent header aspects. This step helps to extract less frequent aspects such as "charging" (sharzhdehi), "battery" (batrish), "body" (badane), "design" (tarrahi), "dimensions" (andaze) and "body material"(jens badaneh) . In a method, we can say that a filter is applied to all candidate aspects, to achieve better performance. The most important feature of the proposed method is that the statistical distribution of the main aspects is extracted on the vector space of the words, from which all the aspects can be extracted with high accuracy.

The dataset used in this study consists of all raw reviews regarding all cellphones, tablets, and laptops available in Digikala website. Digikala is a well-known online shopping website in Iran which offers a variety of products such as digital commodities like tablets, laptops, and cellphones. The final amount of the collected opinions is approximately 450,000 reviews, which forms approximately one million sentences.

It is to be noted that, this step which included the extraction of final aspects, was evaluated with precision, recall criteria and F1. The value of the precision and recall criteria for the final aspects was 0.691 and 0.862, respectively.

B. Aspect – Sentiment Classification

Hitherto, the sentiment and aspect words in a sentence have been extracted. At this point, we will find out what aspect is the target of a sentiment word. If a sentence has only one sentiment word and one aspect, it is clear that the target of the sentiment word is the same one aspect. The main challenge is when we have more than one sentiment and aspect word in one sentence.

There are two general methods to do this step. One is the supervised method and the other is the lexical-based method. Lexicon-based methods often use the distance criterion for aspect-sentiment classification.

This step of the problem is not very challenging. The challenges are related to natural language processing issues, which are outside the scope of this research, and the issue of sentiment analysis. There are a number of these challenges, such as determining the reference of pronouns, dealing with negative phrases in a sentence, and clarifying the ambiguity of the meaning of words. The significant point is that these issues have not yet been resolved in the Persian language.

The proposed method of this research is that for each word of comment, the closest aspect is selected as the target of that sentiment word. Of course, there is a difference that the priority is given between the previous and the next word. As we know, in Persian, the adjective or sentiment word usually comes after the aspect word. Priority refers to the fact that if the distance between two aspects from a word is the same, the aspect before the word is chosen as the target of that comment. The accuracy of this method is 87.1%.

The proposed approach is implemented as a tool called PerSA Tool (Persian Sentiment Analysis Tool) and is published as opensource software. The source codes of all the stages mentioned in the proposed method are accessible here [22].

IV. Analysis and Evaluation

The proposed solution was fully described in the previous section. The main purpose of the method presented in this paper is to extract and cluster aspects as well as aspect-sentiment classification. The word aspect refers to the target of the sentiment or feature about which the opinion is given. This research is being evaluated using the SentiPers labeled dataset [17]. The important point is that the training data was completely different from the evaluation data. The training dataset has been complete without labels.

There are several parameters at different steps of this system. Part of this chapter is devoted to experimentation and the reasons for choosing the values specified for these parameters. These parameters are determined using some labeled data as a validation set. This validation set has also been part of the SentiPers dataset, which is different from evaluation data.

Precision and recall criteria are used for evaluation. The reason for choosing these two criteria is that True Negative cases are not important in them. True Negative cases are the words that are neither considered in the evaluation data nor the output of the proposed system as a word of sentiment or aspect. The involvement of True Negative in the assessment increases the unrealistic measure of accuracy.

A. Determining the main parameters of the problem

The accuracy, recall and F1 criteria are used to determine the main parameters of the problem. The best value for each threshold is the value at which the F1 criterion has the highest value. There are two main parameters in the proposed method.

The first parameter is a threshold that if the cosine similarity of the candidate aspect and one of the leading aspects is higher than this threshold, the candidate aspect is extracted as a final aspect.

The second parameter is a threshold that if the cosine similarity of each word and one of the leading aspects is higher than this threshold, that word will be extracted as the final aspect. The second parameter is added to the system because the word may actually be an aspect, but was not found in candidate aspect is extraction step. The following are the experiments and results for determining each of these parameters.

B. Thresholds of similarity cosine to extract final aspects

In Figure 4-1, the values of accuracy, recall, and F1 can be seen for different values of the first parameter.

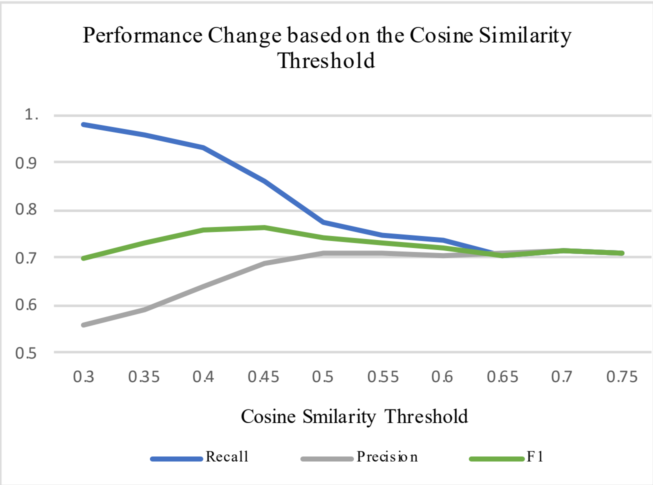

Figure. 4.1. Changes in accuracy according to the similarity threshold of cosine in candidate perspectives

It is clear that the lower the threshold value, the number of final extracted aspects will be greater. A great number of extracted aspects will cause high recall and low precision. On the other hand, the higher the threshold value results in the lower the recall rate and the higher precision. Using the F1 criterion, a balance can be struck between precision and recall. As can be seen in the figure above, the maximum F1 value is when the threshold is 0.45.

The first parameter alone is not enough. The word may be an aspect, but it has not been found at the step of extracting candidate aspects. The second parameter is the threshold, which, if the cosine similarity of each word in the text and one of the main aspects is higher than the threshold, will be extracted as the final aspect.

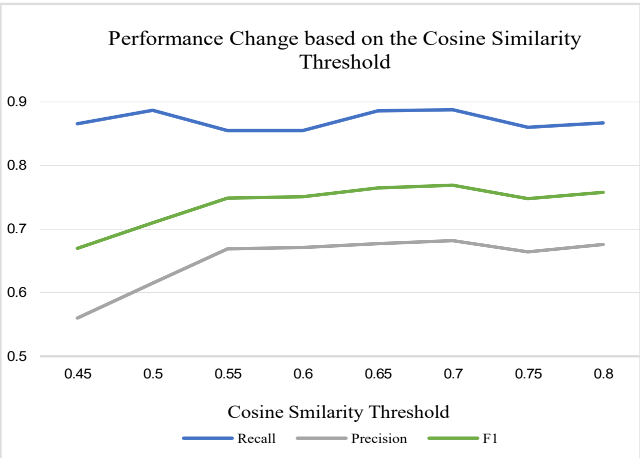

Figure. 4.2. Accuracy changes according to the cosine similarity threshold for all words

As observed in Figure 4-2, the second parameter does not have a major effect on precision. The reason for this, as mentioned earlier, is that this parameter is to complete the previous parameter and has less effect on the accuracy and recall criteria. The application of this parameter is to find words that are not part of the candidate aspect. It is observed that the F1 criterion has the highest value when the threshold is set to 0.7.

Hitherto in this section, the main parameters used in problem-solving steps, and the reasons for their value have been discussed. The proposed method will be evaluated as follows.

V. EVALUATION OF THE PROPOSED METHODS

The value of the main parameters was determined in the former section. The proposed method is evaluated in this section. To make the proposed method more efficient, a part of the SentiPers dataset that is not used to determine the parameters is selected as the evaluation data. It should also be reminded that the learning, validation, and evaluation sets are completely discrete. Also, in this study, no labeled data was used to learn the models and solve the problem.

A. Evaluation the step of extracting and clustering aspects

The step of extracting the final aspects consists of two sub-steps. The first sub-step is to extract candidate aspects using the adjective and descriptive rules. The second sub-step is to extract the final aspects using clustering methods and using cosine similarity criterion. The second step also plays the role of filtering the candidate aspects. Regarding a large number of extracted candidate aspects, these aspects need to be filtered. Table 1 shows the values of accuracy, recall, and F1 for these two steps or two methods.

TABLE 1. Criteria for accuracy, recall, and F1 for aspect extraction

F1	**Recall**	**Precision**	
0.585	0.867	0.442	**Using The Adjective and Descriptive Rules**
0.766	0.862	0.691	**Clustering Methods + Cosine Similarity**

As observed, the aspect extraction method, using adjectives and descriptive rules, has a high recall and low precision. The reason for this is that the number of aspects extracted at this step is very large, which increases recall and decreases precision. In conclusion in this study, the F1 criterion has been increased using clustering methods and these large candidate aspects. This is done by using Word2Vec model vectors and the cosine similarity criterion.

The main advantage of the proposed method is that the aspects are extracted completely unsupervised. The main reason for the emphasis on being unsupervised is that the proposed method does not need or depend on the training data or any other prior information. Eventually, the objective was to make the proposed system independent of the domain. As mentioned in Chapter 3, if we assume that we have the main aspects or features of a product, we can easily extract the other features of the desired product by cosine similarity criterion. But this system does not depend on the given domain and main aspects.

B. Comparison with other researches

Very few researches have been done on the issue of sentiment analysis in Persian. The few works have been done either at the document level or used supervised methods. According to our research, a small number of articles in Persian have solved the step of extracting and categorizing aspects using unsupervised methods. The results obtained from this article and the comparison of our proposed method with the method of Razavi et al. will be presented in this section. For a better comparison, an average of the results provided for laptops, mobile, and TV devices is taken from the Razavi et al [23].

In the following table, the values of accuracy, recall, and F1 are provided for the proposed method and also the method presented in the Razavi et al.

TABLE 2. Comparison of accuracy, recall and F-1 criteria for this research and also, Razavi et al. research

F1	Recall	Precision	
0.741	0.679	0.819	**Razavi et al. method**
0.766	0.862	0.691	**Our Proposed Method**

One of the problems of the proposed method of Razavi deals with obtaining initial aspects or candidate aspects from the method of frequently used nouns. This method has no proper efficiency. This is because many nouns or noun phrases are mistakenly entered in the prospective candidates. As mentioned in the chapter on the proposed method, this article uses the relationship of the adjective between the words aspect and sentiment for the initial aspects.

Another advantage of the proposed method is that it can be extracted from less frequent aspects, which has dramatically increased recall. Of course, accuracy decreases for some value, but the F1 efficiency criterion has increased in general. This is done using the main aspects and a threshold described in the previous chapter.

Also, it is better to convert the candidate aspects into 10 to 15 clusters in order to make the final clustering better observable, understandable, and concise for users. This case has been ignored in Razavi et al. method.

It is clear that the effectiveness of the proposed system in this study will not be as good as supervised methods. However, in this study, we indicated that it is possible to attain acceptable efficiency without using labeled data. Therefore, the main advantage of the proposed method in this study is that it does not depend on a specific scope. Also, in recent years, due to the large volume of raw text data, all researches have tended to apply unsupervised methods. Many studies have also performed well by combining unsupervised and supervised methods. This study also indicates a good method to apply unsupervised methods to solve the problem of sentiment analysis in Persian.

In this chapter, different steps of the proposed system are studied and evaluated, and in the next chapter, the conclusion of this research will be presented in brief. The advantages and disadvantages of the proposed system are also described. There are also suggestions for some future research works.

VI. CONCLUSION

In this paper, we present a proposed method for solving the problem of sentiment analysis using unsupervised methods. The proposed method is a combination of rule-based methods, neural networks, word-embedding-based models, and clustering methods.

Higher initial data results in more accurate vector space and makes the proposed method more efficient as well. The main reason for the inefficiency of the proposed method of this research compared to supervised methods is that labeled data is used in the supervised method.

This labeled data has a large number of aspects given as input to the supervised system. Nevertheless, the proposed system does not use any labeled data about the work scope. As a result, the efficiency of the supervised method in aspect extraction has been higher. However, it is shown that it is applicable to achieve great results without the use of any special supervision. The combination of the proposed method with supervised methods and obtaining better aspects and candidates can be achieved with very high efficiency.

ACKNOWLEDGMENT

The authors would like to thank Navid Alipour for comments that greatly improved the manuscript.

REFERENCES

[1] R. Akhoundzade and K. H. Devin, "Persian Sentiment Lexicon Expansion Using Unsupervised Learning Methods," 2019 9th International Conference on Computer and Knowledge Engineering (ICCKE), Mashhad, Iran, 2019, pp. 461-465, doi: 10.1109/ICCKE48569.2019.8964692.

[2] Liu B. Sentiment analysis and opinion mining. Synthesis lectures on human language technologies. 2012 pp:1-67.

[3] McGlohon M, Glance NS, Reiter Z. Star Quality: Aggregating Reviews to Rank Products and Merchants. In Proceedings of International AAAI conference on Web and Social Media (ICWSM-17) 2010. pp. 1-8.

[4] Asur S, Huberman B. Predicting the future with social media. In Proceedings of Web Intelligence and Intelligent Agent Technology, 2010 IEEE/WIC/ACM International Conference on 2010 Vol. 1, pp. 492-499.

[5] Joshi M, Das D, Gimpel K, Smith NA. Movie reviews and revenues: An experiment in text regression. In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics 2010 pp. 293-296.

[6] Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market. In Proceedings of Journal of Computational Science 2011. Vol. 1, pp. 17-38.

[7] Kessler JS, Nicolov N. Targeting Sentiment Expressions through Supervised Ranking of Linguistic Configurations. In International AAAI conference on Web and Social Media (ICWSM) 2009. pp. 67-78.

[8] Zhuang L, Jing F, Zhu XY. Movie review mining and summarization. In Proceedings of the 15th ACM international conference on Information and knowledge management 2006. pp. 43-50.

[9] Somasundaran S, Namata G, Getoor L, Wiebe J. Opinion graphs for polarity and discourse classification. In Proceedings of the 2009 Workshop on Graph-based Methods for Natural Language Processing 2009. pp. 66-74.

[10] Qiu G, Liu B, Bu J, Chen C. Opinion word expansion and target extraction through double propagation. In Proceedings of the Computational linguistics 37 2011. No. 1, pp. 9-27.

[11] Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems 2013. pp. 3111-3119.

[12] Pavlopoulos J, Androutsopoulos I. Aspect term extraction for sentiment analysis: New datasets, new evaluation measures and an improved unsupervised method. In Proceedings of the 5th Workshop on Language Analysis for Social Media. 2014 Apr 26:44-52.

[13] Hu M, Liu B. Mining and summarizing customer reviews. In Proceedings of the tenth Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD) 2004. pp. 168-177.

[14] Xue B, Fu C, Shaobin Z. A study on sentiment computing and classification of sina weibo with word2vec. In 2014 IEEE International Congress on Big Data 2014. pp. .358-363

[15] Deokar MS. Text Documents clustering using K Means Algorithm. In International Journal of Technology and Engineering Science (IJTES) 2013. Vol 1(4), pp. 282- .286

[16] Jiang M, Zhang Z, Lan M. ECNU at SemEval-2016 Task 5: Extracting Effective Features from Relevant Fragments in Sentence for Aspect-Based Sentiment Analysis in Reviews. In Proceedings of Semantic Evaluation. 2016. pp. 361-366.

[17] Hosseini, P., Ramaki, A.A., Maleki, H., Anvari, M. and Mirroshandel, S.A., 2018. SentiPers: a sentiment analysis corpus for Persian. arXiv preprint arXiv:1801.07737.

[18] Jamadi Khiabani P, Basiri ME, Rastegari H. An improved evidence-based aggregation method for sentiment analysis. Journal of Information Science. 2020 Jun;46(3):340-60.

[19] Dastgheib MB, Koleini S, Rasti F. The application of Deep Learning in Persian Documents Sentiment Analysis. International Journal of Information Science and Management (IJISM). 2020 Feb 9;18(1):1-5.

[20] Dashtipour K, Gogate M, Li J, Jiang F, Kong B, Hussain A. A hybrid Persian sentiment analysis framework: Integrating dependency grammar based rules and deep neural networks. Neurocomputing. 2020 Mar 7;380:1-0.

[21] Shams M, Khoshavi N, Baraani-Dastjerdi A. LISA: Language-Independent Method for Aspect-Based Sentiment Analysis. IEEE Access. 2020 Feb 12;8:31034-44.

[22] Https://github.com/kurosh211/PerSA.git

[23] Razavi, Aref & Asadpour, Masoud. (2017). Word embedding-based approach to aspect detection for aspect-based summarization of persian customer reviews. 1-10. 10.1145/3109761.3158403.