ONOMÁZEIN
Journal of linguistics, philology and translation

PONTIFICIA UNIVERSIDAD
CATÓLICA DE CHILE
FACULTAD DE LETRAS

The role of corpus linguistics in sentiment analysis of Persian texts, case study: a Farsi news agency website

Mohammad Heidarypur
Ferdowsi University of Mashhad
Iran

Mohammad Reza Pahlavannezhad
Ferdowsi University of Mashhad
Iran

Mohsen Kahani
Ferdowsi University of Mashhad
Iran

ONOMÁZEIN 60 (June 2023): 106-121
DOI: 10.7764/onomazein.60.07
ISSN: 0718-5758

Mohammad Heidarypur: Linguistics Department, Faculty of Letters, Ferdowsi University of Mashhad, Iran. | E-mail: mohammad.yours@gmail.com
Mohammad Reza Pahlavannezhad: Linguistics Department, Faculty of Letters, Ferdowsi University of Mashhad, Iran
Mohsen Kahani: Linguistics Department, Faculty of Letters, Ferdowsi University of Mashhad, Iran

Received: February, 2020
Accepted: May, 2020

60
June
2023

Abstract

The current article aims to improve a linguistic model in the sentiment analysis of a Persian news agency website. After investigating many computational problems and shortages in the field of computational linguistics, we could see that the main problem of computational linguists could be found in linguistics, not in computational sciences. Presenting a model can lead to the management of uncertainty of semantic and sentiment analysis of Persian documents. The integration of systems that operate in the field can result in considerable developmental growth in smart systems of sentiment analysis of the Persian language in a way that could reduce complexities in the Persian language. Moreover, the presence of a comprehensive model can facilitate the generation of smart systems of sentiment analysis in text mining. First, we collected existing models for text mining of sentiment analysis and tried to suggest a model as a general principle. The obtained model will help for information management and planning of text mining systems in computational linguistics and shows the shortages of Persian language natural processing in line with the automation of sentiment analysis.

Keywords: sentiment analysis; text mining; corpus linguistics; Persian language; computational linguistics.

1. Introduction

Sentiment analysis is generally a computational study of evaluable expressions in a text. The main goal of sentiment analysis is finding specific opinions of individuals in a text and classifying these opinions. In sentiment analysis, researchers have mainly focused on two problems: detecting whether the text is subjective or objective, and determining whether the text is positive or negative. This level of attention to research and subjective evaluation of opinions in the past few years has coincided with increased online commenting on social media, which has led to the intertwining of this science with computational linguistics and computational sciences. There are several complexities on this path, some of which will be addressed in the present study.

The valuable role of linguists, specifically phonologists, morphologists, grammarians, and semanticists, in languages associated with the translation process stems from the obvious truth: the biggest problems encountered in computer translation are not computer-based but rather language-based (Esmaeil, Hossein, & Fereshteh, 2014). One of the challenges addressed in the present research are the rules existing in universal linguistics, which can disambiguate sentiment analysis in a comment text or commentary on website documents. The author of an opinion or text document is a person who produces a text from electronic content and needs a model to implement rules in natural language processing to analyze the sentiment of the machine-generated text.

There are a variety of topics for reviewing electronic texts on websites, including:

- thematic classification of texts;
- comments on social media;
- authors, time, and other related content extracted from textual documents.

Another challenge are words and their assessment; in this respect, ontology was introduced as a method to retrieve a better meaning from the text.

1.1. Hypothesis and statement of the problem

The hypothesis of natural language processing in an intelligent inferential system is established by identifying a better pattern. Since contextual documents generate written sentences with the natural language of people, it is necessary to follow up on the issues in the area of natural language processing. Some of these issues include explicit or implicit expression of an opinion, adherence to a rule, and determining the language rules in terms of meaning and writing. However, analysis of positive or negative sentiment has been limited in major natural language processing issues (Appel, 2015).

2. Similar works and methods

In a research by Vinodhini & Chandrasekaran (2012) entitled "Sentiment analysis and text mining", the aim was to evaluate the techniques, issues, and methods in semantic analysis. Another goal of the mentioned study was to assess different data sources such as weblogs, sites, datasets, and microblogs. The study included semantic classification topics. In terms of semantic analysis, a literature review revealed two types of techniques based on machine learning and analytics. In machine learning, which is based on classification, two document sets of education and test are ordered. In research on the separation of comparative Persian sentences, machine learning methods (Hossein, Salman, & Saeedeh Sadat) were used to propose a technique to separate comparative sentences, where the lexical and structural features of comparative sentences were expressed. Persian sentences and labels of sentences in the structural and lexical feature extraction unit provide the feature vector to the learning system.

In a research on the quantization of sentiment orientation of Persian text comments of customers on product features on the web, Samaneh & Mohammad Javad extracted key features and expressions and determined sentiment orientation at the levels of words and phrases, critical text sentences, and comment text. To this end, a model was proposed to provide critical text files after selecting the desired goods and collecting opinions. Afterward, the mentioned scholars carried out the pre-processing of comments and unified the writing system and orthography of the texts. This process was performed to find the features of goods based on their characteristics and opinions expressed by customers. The detected features were presented to a computer system, and two modules of creating dictionaries with positive and negative polarity traits and establishing dictionaries for product features extracted the comment patterns by labeling part of speech and labeling features and traits in comments. In the next stage, all comments were extracted based on the determined patterns, and the polarity of each feature was determined based on the pattern. In the end, polarized quantization of all comments was carried out in a conclusion, which showed the visualization of results.

In another study, Xuan et al. (2012) determined the subjectivity of texts using a training set that contained sentences labeled subjective or objective. A set of syntactic patterns was also applied to extract features from texts. In the first stage of determining subjectivity, features were extracted using syntactic patterns, followed by learning with the help of a support-vector machine (SVM) algorithm. In 2011, Chen et al. estimated the semantic indicators of SO-PML. SO-LSA after labeling parts of speech. In the next stage, the features were entered into the nervous systems as input to classify the polarity of sentences. These semantic indicators help the semantic similarity measures between words, and some of the most popular semantic similarity measures include PMI and LSA measures.

In 2013, Li and Tsai proposed a fuzzy-based method to classify ambiguous texts. The suggested model could be used to turn texts into a more abstract form of concepts, which reduced the impact of ambiguous expressions. In addition, Cho et al. (2014) determined the orientation of texts that encompassed diverse ranges of user comments, such as smartphones, movies, and books using sentiment dictionaries. The model proposed for multi-domain analysis is a data-based technique that adjusts sentiment dictionaries to different domains. In research by Ott et al. (2013), a set of opinion spam was created for comments with negative orientation. In the study above, standard techniques of text classification and multiword features were used at the highest level of accuracy to determine whether an opinion is spam or not. Cambria et al. (2014) proposed dictionaries or knowledge bases containing concepts with sentiments related to words or phrases. Other studies have expanded the existing knowledge bases or have used them for sentiment analysis. For instance, Poria et al. (2013) suggested a method to allocate sentiments to concepts existing in the SenticNet knowledge base.

In another study by Parlar et al. (2019), the goals are to analyze the effects of data pre-processing methods for sentiment analysis and determine which of these pre-processing methods (and their combinations) are effective for English as well as for an agglutinative language like Turkish. We also try to answer the research question of whether there are any differences between agglutinative and non-agglutinative languages in terms of pre-processing methods for sentiment analysis. We find that the performance results for the English reviews are generally higher than those for the Turkish reviews due to the differences between the two languages in terms of vocabulary, writing styles, and agglutinative properties of the Turkish language.

According to the study by Mokhosi et al. (2019), aspect-level sentiment analysis has drawn growing attention in recent years, with the higher performance achieved through the attention mechanism. Despite this, previous research does not consider some human psychological evidence relating to language interpretation. This results in attention being paid to less significant words especially when the aspect word is far from the relevant context word or when an important context word is found at the end of a long sentence. We design a novel model using word significance to direct attention towards the most significant words, with novelty decay and incremental interpretation factors working together as an alternative for position-based models. This study demonstrates the relevance of word significance in aspect-level sentiment analysis, which may be applicable in guiding attention to the most significant items in attention-based machine learning tasks.

In another study by Gangrade et al. (2019), sentiments are classified based on Thayer's model, which is psychologically defined, unlike the polarity classification used in opinion mining. In this paper, as a method for classifying the sentiments, sentiment categories are proposed by extracting sentiment keywords for major sentiments by using hashtags, which are essential elements of Instagram. By applying sentiment categories to user posts, sentiments can be determined through the similarity measurement between the sentiment adjective candidates and the sentiment keywords.

According to another study, by Karamitsos et al. (2019), the availability and advancements of cloud computing service models such as IaaS, SaaS, and PaaS, introducing on-demand self-service, auto-scaling, easy maintenance, and pay-as-you-go, have dramatically transformed the way organizations design and operate their data centers. However, some organizations still have many concerns, like security, governance, lack of expertise, and migration. The purpose of this paper is to discuss the cloud computing customers' opinions, feedback, attitudes, and emotions towards cloud computing services using sentiment analysis. The associated aim is to help people and organizations to understand the benefits and challenges of cloud services from the general public's perspective view as well as opinions about existing cloud providers, focusing on three main cloud providers: Azure, Amazon Web Services (AWS), and Google Cloud. The methodology used in this paper is based on sentiment analysis applied to the tweets that were extracted from the social media platform (Twitter) via its search API.

3. Materials and methods

In the present study, we used the news on the website of Alef News Agency. Some of the features of the news on this website include: 1) related news features: a list of news related to the observed news is retrieved using the selection algorithm; 2) keyword feature: titles, labels, and descriptions of news form keywords that reflect the conceptual meaning of news; 3) feature of opinions: users can provide feedback on the news they see, and other users can read the feedback.

In [1], four textual features were shown: titles, labels, descriptions, and opinions were effective in improving the classification accuracy of social media. Therefore, the features were applied to facilitate the process of determining the semantic load of news.

After preprocessing of news and extraction of four textual features, the news was classified into sentiment categories using the approach of sentiment dictionary of unsupervised machine learning (UML) and supervised machine learning algorithm (SML). Afterward, the two methods were combined to develop a new group model for sentiment classification.

In the study, textual features of news on the website of Alef News Agency (news title, tag, descriptions, and opinions) were collected as research data. Moreover, approaches of ML (supervised learning) and sentiment dictionary (unsupervised learning) were deployed to classify news into proper sentiment categories.

4. Implementation

The news data used in this study consisted of four textual sections, and the news was extracted from the website of Alef News Agency through the following stages:

Steps: News with specific features are first identified using relevant topic terms and then approved by experts and classified into the correct sentiment classification. In this test, a Telegram robot was used to fetch the dataset of the website of Alef News Agency, features of which were included in opinions, descriptions, hashtags, and titles. Two experts working in the field of data mining and sentiment analysis classified the news into five categories of sentiments. If consensus was reached on the news, the news would be entitled "sentiment based on satisfaction of parties" during the naming process. On the other hand, the lack of consensus would lead to adding the news to the elimination list. In total, 1217 news were maintained with 200 news in each sentiment category.

In the research, we applied the 10-fold cross-validation method, where the original sample was randomly partitioned into 10 subsamples after performing the cross-validation process 10 times. Also, we deployed a non-parametric one-sample test to recognize the main differences between ten times the execution of the data. In this respect, all results showed a lack of significant difference between the implementation times ($P \geq 0.05$). After completing the classification, some classifiers were recognized using an SVM and were compared to those identified by using other classifications. All machine-learning implementations were obtained through classifiers on Python software.

Descriptions of laboratory methods accepted in the study are presented in Table 1. We considered three indexes of accuracy, FI value, and AUC value to assess the performance of experiments.

TABLE 1

Experimental methods

NUMBER	TEST METHOD	DESCRIPTION
1	M1: Keyword baseline	ALM provided an unsupervised sentiment classification approach where keywords were labeled using a sentiment dictionary.
2	M2: Sentiment dictionary approach	An unsupervised sentiment classification approach was used in this study to measure the correlation between the news and numerous sentimental features by estimating the PMI correlation between sentimental expressions in the sentiment dictionary and expressions observed in the news on the website of Alef News Agency.
3	M3: Simple bias ML with three methods of TFIDF	Feature vectors were first generated using the three methods of TFIDF. Afterward, the sample bias algorithm was applied to classify the news.

4	M4: J48 decision tree with three methods of TFIDF	The feature vectors were first generated by three methods of TFIDF. Afterward, the J48 decision tree was used to classify the news.
5	M5: SVM ML with three methods of TFIDF	Feature vectors were first generated by methods of TFIDF. Afterward, the SVM algorithm was applied to classify the news.
6	M6-1, 6-2: Group model 1	A classification approach has been designed in the research to classify the news having the most level of sentiments. It was estimated by multiplying the probability of sentiments obtained by the ML method by the probability obtained in the method of sentiment dictionary (M2).
7	M7-1, M7-2, M7-3, M7-4, M7-5, M7-6: Group model 2	Group model 2 was used in two stages. In the first stage, probabilities of multiple categories were generated using ML and sentiment dictionary approaches. In the second stage, probabilities were used as feature amounts, and the main sentiment category was selected as a label. In the next stage, SVM and simple bias were deployed to practice the learning models.

4.1. Results of tests

4.1.1. Approach of sentiment dictionary

Considering the sentiment dictionary approaches used in unsupervised learning, the sentiment dictionary approach regarded in the present study was compared to the keyword baseline labeling approach introduced by Alm et al. (9). According to Table 2, while PMI generated higher accuracy compared to keyword baseline labeling, it is lower than keyword baseline labeling considering FI. However, lower accuracy rates are obtained from sentiment dictionary approaches. Furthermore, we tested the important statistical differences between the two methods using the McNemar test, the results of which were indicative of considerable differences in terms of the accuracy of techniques.

TABLE 2

Accuracy, FI value, and AUC of sentiment dictionary methods

NUMBER	SENTIMENT DICTIONARY METHODS	ACCURACY	F1	AUC
1	Keyword baseline (M1)	37.01%	0.640	0.476
2	PMI (M2)	42.01%	0.531	0.487

4.1.2. ML approaches

About the ML approaches used in the supervised learning method, the news was defined with proper sentiment classifications and was turned into a vector applying three vector labeling methods of TFIDF. Afterward, we used the naïve Bayes, J48, and SVM algorithms to classify news based on the classification models. According to Table 3, the accuracy rate obtained from supervised learning was significantly higher than what can be obtained by unsupervised learning. Moreover, ML methods, which are specifically designed for analysis of news on the website of Alef News Agency (i.e., methods 2 and 3 of TFIDF), had a more efficient performance in most cases, compared to the old TFIDF method.

TABLE 3

Accuracy, F1 value, and AUC of sentiment dictionary methods

NUMBER	MACHINE LEARNING		SIMPLE BIAS (M3)	J48 DECISION-MAKING TREE (M4)	SVM (M5)
1	TFIDF1 (baseline)	Accuracy F1 AUC	65.82% 0.682 0.899	72.56% 0.782 0.896	76.42% 0.815 0.965
2	TFIDF2	Accuracy F1 AUC	82.15% 0.828 0.953	85.94% 0.848 0.923	65.13% 0.649 0.929
3	TFIDF3	Accuracy F1 AUC	75.10% 0.755 0.940	86.20% 0.863 0.942	85.13% 0.878 0.982

4.1.3. Group model

Group model 1 combines the M3 and M5 methods with the approach of sentiment dictionary (M2). The probability values were determined for different product combinations by the sentiment dictionary or ML approaches. According to Table 4, the accuracy rate of group model 1 was lower than the rate obtained by independent use of ML, showing the lack of ability of the model to improve the accuracy rate of ML.

In the group model, the probability values of a combination of the dictionary approach and the sentiment dictionary method were regarded as new values, whereas the main categories of sentiment were applied as labels. Also, logistic regression, SVM, and naïve Bayes algorithms were subsequently used for the education of learning models. According to Table 5, three combinations of group model 2 (e.g., M7-4, M7-2, and M7-6) yielded much better results, compared to the results obtained from the independent use of ML or sentiment dictionary approach. A common point among all three methods is their use in the first stage of SVM. From the average results, optimal results were obtained when using SVM and sentiment dictionary approaches in the first stage and adopting the logistic regression algorithm in the second stage (M7-2). Compared to the old ML method, the mentioned model has a 10% higher accuracy, which shows that the main advantage has been created through the integration of supervised and unsupervised approaches.

TABLE 4

Accuracy, F1 value, and AUC of group model 1

NUMBER	MACHINE LEARNING		M6-1 (M2+ M3)	M6-2 (M2+ M5)	M3	M5
1	TFIDF1 (baseline)	Accuracy F1 AUC	56.45% 0.798 0.930	63.27% 0.752 0.939	65.82% 0.682 0.899	76.42% 0.815 0.965
2	TFIDF2	Accuracy F1 AUC	82.15% 0.832 0.952	68.09% 0.458 0.831	82.15% 0.828 0.958	65.13% 0.649 0.889
3	TFIDF3	Accuracy F1 AUC	59.57% 0.766 0.910	73.04% 0.793 0.852	75.10% 0.755 0.940	85.13% 0.878 0.982

TABLE 5

Accuracy, F1 value, and AUC in group model 2

		M7-1 (M2 + M3)	M7-2 (M2 + M5)	M7-3 (M2 + M3)	M7-4 (M2 + M5) SVM	M7-5 (M2 + M3)	M7-6 (M2 + M5)	M3	M5
		Logistic	Logistic	SVM	SVM	Naive Bayes	Naive Bayes		
TFIDF1	**ACCURACY**	72.64%	83.65%	70.74%	79.70%	71.89%	80.03%	65.82%	76.42%
	F1	0.729	0.840	0.715	0.723	0.811	0.804	0.682	0.815
	AUC	0.923	0.970	0.892	0.904	0.963	0.955	0.899	0.965
TFIDF2	**ACCURACY**	81.99%	80.26%	82.24%	79.44%	81.90%	72.53%	82.15%	65.13%
	F1	0.820	0.802	0.625	0.706	0.727	0.821	0.828	0.649
	AUC	0.946	0.958	0.856	0.932	0.935	0.922	0.958	0.889
TFIDF3	**ACCURACY**	77.16%	89.89%	75.43%	88.08%	76.17%	88.66%	75.10%	85.13%
	F1	0.773	0.899	0.759	0.892	0.764	0.889	0.755	0.878
	AUC	0.938	0.986	0.916	0.984	0.972	0.978	0.940	0.982
AVERAGE	**ACCURACY**	77.26%	84.60%	76.13%	82.40%	76.65%	80.40%	74.35%	75.56%
	F1	0.774	0.847	0.700	0.774	0.767	0.838	0.755	0.781
	AUC	0.936	0.971	0.888	0.940	0.665	0.952	0.932	0.945

Moreover, we tested the significant statistical differences between the pair of methods by a two-way comparison, where a non-parametric test (McNemar) was applied for all 11 methods (from M3 to M6-7) to determine the significant differences in this regard. In the research, the binary P-values were reported at the significance level of P<0.05. Due to the use of three data formats (TFIDF1, TFIDF2, and TFIDF3), each cell in Table 6 contained three values showing the p-values for accuracy in formats of TFIDF1, TFIDF2, and TFIDF3 (presented by this order). For instance, the values of (M3, M7-1) cell were *0.000, 0.590, and 0.259, which were the corresponding values of P for pair of (M3, M7-1) for TFIDF1, TFIDF2, and TFIDF3. In this context, "*" showed the statistical significance of the results.

TABLE 6

Significant results for comparison of accuracy

	M6-1	M6-2	M7-1	M7-2	M7-3	M7-4	M7-5	M7-6	M3	M4
M6-2	0.000*									
	0.000*									
	0.000*									
M7-1	0.000*	0.000*								
	0.917	0.000*								
	0.000*	0.008*								
M7-2	0.000*	0.000*	0.000*							
	0.519	0.000*	0.631							
	0.000*	0.000*	0.000*							
M7-3	0.000*	0.000*	0.007*	0.000*						
	1.000	0.000*	0.874	0.560						
	0.000*	0.154	0.024*	0.000*						
M7-4	0.000*	0.000*	0.000*	0.000*	0.000*					
	0.000*	0.000*	0.000*	0.000*	0.000*					
	0.000*	0.000*	0.000*	0.006*	0.000*					
M7-5	0.000*	0.000*	0.374	0.000*	0.000*	0.000*				
	0.876	0.000*	1.000	0.672	0.250	0.000*				
	0.000*	0.043*	0.182	0.000*	0.185	0.000*				
M7-6	0.000*	0.000*	0.000*	0.030*	0.000*	0.676	0.000*			
	0.014*	0.002*	0.000*	0.000*	0.000*	0.000*	0.000*			

	0.000*	0.000*	0.000*	0.050*	0.000*	0.188	0.000*			
M3	0.000*	0.000*	0.000*	0.000*	0.013*	0.000*	0.000*	0.000*		
	0.709	0.000*	0.590	0.276	0.667	0.000*	0.915	0.000*		
	0.000*	0.254	0.259	0.000*	0.926	0.000*	0.575	0.000*		
M4	0.000*	0.000*	1.000	0.001*	0.000*	0.005*	0.748	0.000*	0.000*	
	0.036*	0.000*	0.025*	0.006*	0.031*	0.000*	0.024*	0.000*	0.002*	
	0.000*	0.000*	0.000*	0.009*	0.000*	0.219	0.000*	0.098	0.000*	
M5	0.000*	0.000*	0.038*	0.000*	0.000*	0.050*	0.000*	0.000*	0.000*	0.026*
	0.000*	0.143	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.021*	0.000*
	0.000*	0.000*	0.000*	0.000*	0.000*	0.041*	0.216	0.013*	0.000*	0.432

Three observations were obtained from Tables 3 to 6 (1). The M7-2 had the most suitable performance, compared to other techniques, when adopting the TFIDF1 format, in a way that all primary P-values were significant in all cells in the column and row of M2-7 (2). Similarly, the M4 yielded considerably better results, compared to other methods, when adopting the TFIDF2 format (3). On the other hand, the M2-7 outperformed all other methods when adopting the TFIDF3 format. Accuracy of 83.65%, 85.94%, and 89.89% was obtained by three combinations of TFIDF1+M2-7, TFIDF2+M4, and TFIDF3+M7-2. While the TFIDF3+M7-2 was the most accurate combination, we re-assessed its significance using the McNemar test. According to Table 7, the combination of TFIDF3+M7-2 had a more efficient performance, compared to the other two combinations. Therefore, the combination of TFIDF3+M7-2 was recognized as the most accurate model. In other words, it is suggested that TFIDF3 be used as data format and M7-2 (first stage: M2+M5, second stage: logistic) be applied as the solution model to achieve the most accurate prediction of accuracy.

TABLE 7

Comparison of three more accurate combinations

	T1M7-2	T2M4
T2M4	0.261	
T3M7-2	0.000*	0.001*

Ultimately, we analyzed the robustness of our algorithms. In this context, robustness is defined as the ability of an algorithm to predict the most difficult prediction class accurately. In other words, the algorithm operates reasonably well in the most unfavorable conditions regardless of the class distribution. In the study, we used the lowest F1 value among the five sentiment classes to measure robustness. The relatively high level of the lowest F1 value was interpreted as a suitable prediction performance of the algorithm even for the sentiment class with the most difficult mode of prediction. Table 8 shows the lowest F1 values for all methods in three data formats. According to the table, M4 and M7-2 were the most robust models, specifically in the use of TFIDF2 and TFIDF1, respectively. However, the mentioned methods had comparable performance when using TFIDF3. Nonetheless, the mean accuracy of M4 (81.56) was lower than the M7-2 (84.60), as shown in Tables 2 and 5. Therefore, while M2-7 was more accurate, the M7-2 and M4 had similar robustness (P-value=0.000 based on the McNemar test).

TABLE 8

Robustness of the algorithm under three data formats

	M3	M4	M5	M6-1	M6-2	M7-1	M7-2	M7-3	M7-4	M7-5	M7-6
TFIDF1	0.543	0.705	0.541	0.000	0.377	0.575	0.714	0.555	0.609	0.578	0.644
TFIDF2	0.711	0.783	0.000	0.670	0.107	0.740	0.620	0.736	0.107	0.731	0.418
TFIDF3	0.627	0.798	0.609	0.211	0.000	0.661	0.796	0.634	0.763	0.647	0.768

5. Conclusion

In the present research, three methods were used in the ML approach to creating feature vectors. According to the results, methods of TFIDF2 and TFIDF3, where feature vectors were constructed and corrected based on the website of Alef News Agency, the classification accuracy was more satisfactory, compared to the traditional method of TFIDF1. The results of comparison with unsupervised learning methods of labeling based on Alm keywords [9] showed that in the sentiment dictionary approach, PMI yielded more accurate results in terms of calculation of sentimental correlation and classification of sentiment features. In addition, the ML approaches and sentiment dictionary were combined to form cumulative models. According to the results, this combination successfully increased the accuracy rate in group model 2 (where probabilities are multiplied), and the results were more satisfactory than what would be obtained from the independent use of the ML approach or sentiment dictionary. The results confirmed that the proposed method effectively facilitated the classification of news on the website of Alef News Agency into proper categories. One of the major drawbacks of the present research was collecting data from the website of Alef News Agency. Therefore, it is suggested that a wider range of data or empirical data be obtained from other news platforms in future studies. Another limitation of the research was the evaluation of only four textual features (e.g., labels, descriptions, and opinions). It is recommended that more features be assessed in future studies (e.g., audio, video, and the view count of news). In the present study, the method applied was a combination of SML and UML approaches. A limitation of SML is its need for an education dataset with pre-determined labels. In other words, the sentiment group of each news must be first defined by experts to use our technique. Unfortunately, manual labeling is often associated with a huge cost of human labor, and it is difficult to obtain a large amount of high-quality educational data using this method.

Obtaining labeled datasets is often hard, costly, and time-consuming since their preparation requires a great effort on the part of a trained human annotator. A suitable alternative approach is the application of semisupervised learning (SSL), which is between supervised and unsupervised paradigms. For better classifications, SSL uses a high amount of unlabeled data along with labeled data. Future studies must consider the use of SSL-based methods to classify news based on sentiments since unlabeled data can be easily collected, need less human effort, and often have higher accuracy.

In addition, evaluation of former studies demonstrated that the effectiveness of the unsupervised learning method was generally lower, compared to the supervised learning approach. Nevertheless, an unfavorable classification accuracy rate was detected when using the sentiment dictionary approach. This issue might be due to insufficient coverage of various items in the sentiment dictionary approach used in the current study.

Developed in 2010, the NRC sentiment dictionary might not completely cover all phrases used by individuals in different age groups. Coverage is a fundamental and difficult research topic in sentiment dictionary approaches. To increase the accuracy of this method, it is recommended that a new dictionary be developed in future studies with an emphasis on common phrases used in comments of users to solve the problem of improper coverage in the current sentiment dictionaries.

Finally, all news considered was limited to just one type of sentiment. Any news that was difficult to judge was simply omitted from the analysis. Thus, in practice, 89% accuracy represents only performance in easy-to-judge news and not all news on the Alef site. In the future, we might attempt to expand the news classification method with multiple sentiment categories.

6. References

Abbasi, Ahmet, Hsinchun Chen & Arab Salem, 2008: “Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums”, *ACM Transactions on Information Systems (TOIS)* 26 (3), 1-34.

ALM, Ebba Cecilia Ovesdotter, 2008: *Affect in Text and Speech*, Illinois: the University of Illinois, Urbana-Champaign.

APPEL, Orestes, Francisco CHICLANA & Jenny CARTER, 2015: "Main Concepts, State of the Art and Future", *Acta Polytechnica Hungarica* 12 (3), 87-108.

CAMBRIA, Erik, 2013: "An introduction to concept-level sentiment analysis" in Félix CASTRO, Alexander GELBUKH & Miguel GONZÁLEZ (eds.): *Advances in Soft Computing and Its Applications*, Berlin/Heidelberg: Springer, 478-483.

FIGUEIREDO, Flavio, Henrique PINTO, Fabiano BELÉM, Jussara ALMEIDA & Marcos GONÇALVES, 2013: "Assessing the quality of textual features in social media", *Information Processing & Management* 49 (1), 222–247.

GANGRADE, Shweta, Nirvishesh SHRIVASTAVA & Jayesh GANGRADE, 2019: "Instagram Sentiment Analysis: Opinion Mining" in *Proceedings of Recent Advances in Interdisciplinary Trends in Engineering & Applications (RAITEA)*.

CHO, Heeryon, Songkuk KIM, Jongseo LEE & Jong-Seok LEE, 2014: "Data-driven integration of multiple sentiment dictionaries for lexicon-based sentiment classification of product reviews", *Knowledge-Based Systems* 71 (1), 61-71.

XUAN, Huong Nguyen Thi, Anh Cuong LE & Le Minh NGUYEN, 2012: "Linguistic Features for Subjectivity classification" in *Proceedings of the International Conference on Asian Language Processing (IALP)*, 17-20.

KARAMITSOS, Iannis, Saeed ALBARHAMI, Charalampos APOSTOLOPOULOS, 2019: "Tweet Sentiment Analysis (TSA) for Cloud Providers Using Classification Algorithms and Latent Semantic Analysis", *Journal of Data Analysis and Information Processing* 7 (4), 276-294.

CHEN, Long-Sheng, Cheng-Hsiang LIU, Hui-Ju CHIU, 2011: "A neural network-based approach for sentiment classification in the blogosphere", *Journal of Informetrics* 5 (2), 313-322.

MOKHOSI, Refuoe, ZhiGuang QIN, Qiao LIU, Casper SHIKALI, 2019: "Incorporating Word Significance into Aspect-Level Sentiment Analysis", *Applied Sciences* 9 (17), 3522.

PARLAR, Tuba, Selma Ayşe ÖZEL, Fei SONG, 2019: "Analysis of data pre-processing methods for the sentiment analysis of reviews", *Computer Science* 20 (1), 123-141.

PORIA, Soujanya, Erik CAMBRIA, Grégoire WINTERSTEIN, Guang-Bin HUANG, 2014: "Sentic patterns: Dependency-based rules for concept-level sentiment analysis", *Knowledge-Based Systems* 69 (1), 45-63.

LI, Sheng-Tun, Fu-Ching TSAI, 2013: "A fuzzy conceptualization model for text mining with application in opinion polarity classification", *Knowledge-Based Systems* 39 (1), 23-33.

VINODHINI, G., & R. M. CHANDRASEKARAN, 2102: "Sentiment Analysis and Opinion Mining: A Survey", *International Journal* 2 (6), 282-292.