An Ensemble Based Classification Approach for Persian Sentiment Analysis

Kia Dashtipour, Cosimo Ieracitano, Francesco Carlo Morabito, Ali Raza, and Amir Hussain

Abstract In recent years, sentiment analysis received a great deal of attention due to the accelerated evolution of the Internet, by which people all around the world share their opinions and comments on different topics such as sport, politics, movies, music and so on. The result is a huge amount of available unstructured information. In order to detect positive or negative subject's sentiment from this kind of data, sentiment analysis technique is widely used. In this context, here, we introduce an ensemble classifier for Persian sentiment analysis using shallow and deep learning algorithms to improve the performance of the state-of-art approaches. Specifically, experimental results show that the proposed ensemble classifier achieved accuracy rate up to 79.68%.

Keywords Persian sentiment analysis · Natural language processing · Deep learning · Ensemble classifier

K. Dashtipour (✉)
Department of Computing Science and Mathematics, University of Stirling, Stirling, UK
e-mail: kd28@cs.stir.ac.uk

C. Ieracitano · F. Carlo Morabito
DICEAM, University Mediterranea of Reggio Calabria, Via Graziella, Feo di Vito, 89060 Reggio Calabria, Italy

A. Raza
Department of Electrical Engineering and Computing, Rochester Institute of Technology, Dubai, UAE

A. Hussain
School of Computing, Edinburgh Napier University, Merchiston Campus, Edinburgh EH10 5DT, UK

A. Esposito et al. (eds.), *Progresses in Artificial Intelligence and Neural Systems*, Smart Innovation, Systems and Technologies 184,
https://doi.org/10.1007/978-981-15-5093-5_20

1 Introduction

Sentiment analysis (SA) is the computational study of peoples behaviours, sentiments, emotions and attitudes through the extraction of significant information from unstructured big data (such as products and services). The rise of social media, blogs and forums allow people to share continuously comments and ideas on different topics. For example, in buying a consumer product, there is no longer any need to ask friends or families, as many reviews and discussions are available online on that specific product. Similarly, for a company or organization, there is not anymore need to conduct surveys to understand public opinions about their services as social media help to reorganize their business. SA uses natural language processing (NLP) to extract useful information such as identifying the polarity of the text from online data [9, 10].
There are different types of opinions available online:

- Explicit opinion: expresses opinion about an entity directly. For example, “I do not like Spider-man movie” من فيلم مرد عنكبوتي دوست نَدارم. .
- Comparative opinions: expresses by comparing one entity by another entity. For example, “The spider-man movie is better than Avengers” بهتر از انتقَام جويّان است فيلم مرد عنكبوتي. .

However, most of the current approaches are available in English and only a few works regard other languages such as Persian (the official language of Iran and Afghanistan that includes more than 80 million speakers). Hence, in order to fill the gap, we propose a novel framework for Persian sentiment analysis. Specifically, as deep learning (DL, [15, 16]) has been achieving impressive results in several real-world problems ([14, 18–21]), we propose an ensemble classifier based on shallow (Support Vector Machines (SVM), Multilayer Perceptron (MLP)) and DL (Convolution Neural Network (CNN)) techniques to detect polarity in Persian sentiment analysis. The hotel reviews dataset gathered from [3] is used to evaluate the proposed approach.

The rest of the paper is organized as follows: Sect. 2 presents related work. Section 3 presents methodology and experimental results. Finally, Sect. 4 concludes this paper.

2 Related Work

In the literature, extensive research was carried out to model novel sentiment analysis models using both shallow and deep learning algorithms. For example, García-Pablos et al. [13] proposed a supervised approach for aspect-based sentiment analysis. The approach was evaluated using different languages such as English, Spanish, French and Dutch. The hotel, restaurants and electronic devices datasets are used to evaluate the performance of the approach. Experimental results showed that the proposed approach (89%) obtained better accuracy as compared with CNN (82.31%) and LSTM (79.25%). Dashtipour et al. [8] proposed a framework for using deep learning classifiers to detect polarity in Persian movie reviews. The proposed deep learning classifiers were compared with shallow MLP. Simulation results reported that DL classifiers such as 1D-CNN (82.86%) showed better performance as compared with MLP (78.49%). Dos Santos et al. [11] proposed an approach to detect polarity in short text using deep learning classifiers. The character embedding used to convert the short text. Experimental results showed the CNN (86.4%) achieved better accuracy as compared with LSTM (85.7%). Poria et al. [30] proposed a novel framework for detecting concept sentiment analysis which merges linguistic patterns and machine learning to identify the polarity of the sentence. Experimental results showed the proposed approach (76.33%) achieved better performance as compared with sentic pattern (70.84%). However, the linguistic patterns are used for English sentiment analysis and it cannot detect polarity in any other language such as Persian. Sohangir et al. [33] proposed a model using deep learning classifiers to detect sentiment for StockTwits. Experimental results demonstrated that the CNN outperformed LSTM, doc2vec and logistic regression, achieving accuracy rate up to 90.93%. Al-Smadi et al. [2] proposed an aspect-based sentiment analysis for Arabic hotel reviews using LSTM and neural network. The character level bidirectional LSTM used for aspect extraction in Arabic hotel reviews. Experimental results demonstrated that the proposed approach (82.7%) achieved better performance as compared with the traditional approach with lexicon (76.4%). Nakayama et al. [25] proposed a method to detect polarity in Japanese hotel reviews. The Yelp hotel reviews used which consist of 157 million reviews. The results are compared with English reviews. The experimental results indicated the English reviews explicitly express the polarity in their comments as compared with Japanese reviews. Poria et al. [29] proposed a parser to breaks texts into words and extract meaningful concept from sentences. There are different rules are developed to identify the polarity of the sentence. The experimental results showed that the proposed approach (86.10%) received better performance as compared with the part-of-speech tag (92.21%). Ozturk et al. [27] proposed a novel method to detect polarity in public opinions towards the Syrian refugee crisis. There are 2,381,297 million tweets are collected in English and Turkish languages. The results indicate the Turkish tweets express more positive opinions towards Syrian refugees than English tweets. On the other hand, most of the English tweets expressed neutral sentiments. However, the proposed method did not use any machine learning classifiers. Li et al. [23] proposed the architecture for sentiment recognizer on the call centre. The proposed architecture used openSMILE to identify the polarity of the sentences. The experimental results indicated the architecture is successful to identify the polarity of the sentence. Minaee et al. [24] presents a model based on ensemble classification of deep learning classifiers to capture the temporal information of the data (English dataset). The experimental results indicated the ensemble classification (90%) achieved better accuracy as compared with CNN (89.3%) and LSTM (89%).

Table 1 Summarized results for different approaches

References	Purpose	Language	Approach	Accuracy
Chen et al. [6]	Detect polarity in Chinese reviews	Chinese	LSTM	78%
Shuai et al. [32]	Detect polarity in Hotel reviews	Chinese	SVM	81.16%
Kirilenko et al. [22]	Detect polarity for Tourism	English	SVM	75.23%
Al-Smadi et al. [1]	Detect sentiment in Arabic reviews	Arabic	Deep Recurrent neural network SVM	78%
Dashtipour et al. [7]	Feature combination for Persian	Persian	SVM	81.24%
Cambria et al. [5]	Develop lexicon for English	English	CNN	94.6%
Dragoni et al. [12]	Detect polarity in multidomain	English	SVM	88.81

Rogers et al. [31] developed a new corpus for Russian language. The data is collected from social media and it was labelled into positive, negative and neutral. After annotation, the data converted using word2vec (fastText) and TF-IDF. The experimental results demonstrated neural net classifier (71.7%) achieved better F-measure as compared with Linear SVM (62.6%) and logistic regression (63.2%). Hazarika et al. [17] proposed a model to detect sarcasm expressed in the text. The hybrid model employed content and context-driven for detecting sarcasm in social media discussions. CNN is trained to detect sarcasm in the sentence. Experimental results indicated discourse features and embedding of the words play important roles to detect sarcasm in the sentence. Peng et al. [28] proposed an approach to detect sentiment in Chinese reviews. The SemEval dataset used to evaluate the performance of the approach. Experimental results showed the proposed approach (75.59%) achieved better accuracy as compared with SVM (66.92%), LSTM (74.63%), BiLSTM (74.15%). In Table 1 some of the sentiment analysis approaches are depicted. However, the none of the aforementioned studies explored the ensemble classifier to detect polarity for Persian sentences. Most of the current approaches use instead ensemble classifier to identify the polarity for English sentences. In contrast, in the present research, we propose a novel ensemble based classification framework for Persian sentiment analysis.

3 Methodology

The proposed methodology includes three main stages: *data pre-processing, feature extraction* and *classification*. Each processing module is described as follows.

3.1 Data Description and Pre-processing

The hotel reviews dataset gathered from [3] is used in this work. consists of 3000 reviews: 1500 positive and 1500 negative. For classification purpose, 60% of data is used as a train set, 30% as a test set and finally and 10% as a validation set. The corpus was pre-processed using the following techniques: *tokenisation*, *normalisation* and *stemming*.

- Tokenisation is used to convert the sentences into words or tokens. For example, فِلم خوب بود (Movie is great), is divided into فیلم, خوب, بود ; or, مر۳۰ (Thanks30) is converted into مرسی.
- Normalisation technique is used to convert these words into their normal forms;
- Stemming is the process of converting words into their roots.

3.2 Feature Extraction

After the data pre-processing, the N-gram (unigram, bigram and trigram) features are extracted. N-grams represent continuous sequences of *n* items in the text. When n = 1 the ngram is called unigram, when n = 2 bigram, when n = 3 trigram and so on. For example, “I like this movie” من این فیلم دوست دَارم . The unigram is من , این , فیلم , دوست , دَارم , “I”, “like”, “this”, “movie”. The bigram features are من فیلم , فیلم دوست , دوست دَارم and trigram features are من فیلم دوست and فیلم دوست دَارم.

3.3 Classification

In order to classify negative and positive reviewers, standard (i.e. SVM, MLP) and deep (CNN) machine learning classifiers are employed. In order to train the classifiers, we used word2vec. Specifically, the sentences are converted into 300-dimensional vectors by using fastText (a Python package) [4].

Convolutional neural network (CNN). The proposed CNN consists of 11 layers (4 convolution layers, 4 max pooling and 3 fully connected layers). Convolution layers have 15 filters sized 1×2 with stride size 1. Each convolution layer is followed by a max polling layer with window size 1×2. The last max pooling layer is followed by a standard MLP with hidden layers size 5000, 500 and 4. In the final layer, softmax activation function is used for classification purpose.

Support Vector Machine. The support vector machines (SVM) is used to find the decision boundary to separate different classes. The Sklearn Python package is used to train the proposed SVM classifier. In addition, the linear kernel is used.

Multilayer Perceptron (MLP). The MLP is a supervised machine learning technique. It typically consists of an input, hidden and output layer. Here, a single hidden layer MLP with 50 hidden units was developed and trained for 100 iterations. A softmax output layer was then used for positive vs negative classification task. In addition, the alpha 0.5 and adaptive learning rate is used.

Ensemble Classifier. Ensemble method consists of employing different classifiers and combining their predictions to train a meta-learning model. The ensemble is typically used to enhance the accuracy of a specific system [26]. In this study, the predictions of the aforementioned classifiers (SVM, MLP, 1D-CNN) are used as input of a linear-SVM based architecture. Figure 1 shows the proposed ensemble classification system. It is to be noted that the parameters and topology of each classifier has been set-up empirically after several simulation experiments.

3.4 Experimental Results

In order to evaluate the performance of the proposed approach, precision, recall, F-measure and accuracy metrics were used:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$F\_measure = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{3}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

where TP denotes true positive, TN is true negative, FP is a false positive, and FN is false negative. The experimental results are shown in Table 2. As can be seen, MLP achieves accuracy of 74.26%, 72.22%, 63.23% when unigram, bigram, trigram were

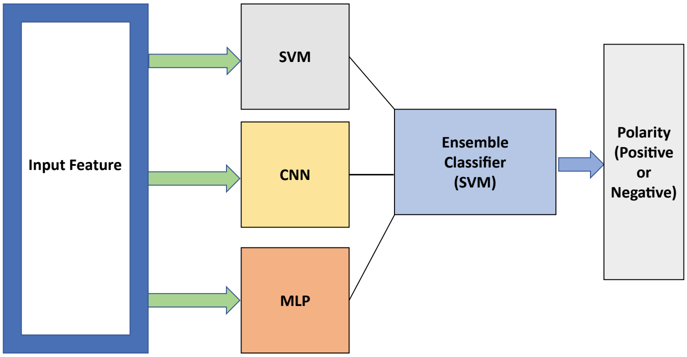

Fig. 1 Proposed ensemble classifier

Table 2 Results of the proposed SVM, MLP, 1D-CNN and ensemble classifier

Method	Input	Precision	Recall	F-measure	Accuracy
SVM	Unigram	0.68	0.64	0.64	68.25
	Bigram	0.71	0.65	0.59	64.92
	Trigram	0.69	0.59	0.45	58.80
MLP	Unigram	0.76	0.74	0.74	74.26
	Bigram	0.74	0.72	0.71	72.22
	Trigram	0.69	0.63	0.56	63.23
1D-CNN	Unigram	0.73	0.80	0.76	78.02
	Bigram	0.71	0.76	0.72	74.25
	Trigram	0.70	0.73	0.71	72.45
SVM (ensemble classifier)	Unigram	0.79	0.78	0.76	79.68
	Bigram	0.80	0.79	0.75	78.18
	Trigram	0.73	0.80	0.76	78.02

used as input respectively. As regards SVM the optimal result was observed with unigram features input (accuracy of 68.25%). The 1D-CNN, instead achieved better classification accuracy, reporting 78.02%. However, the proposed ensemble classifier (based on SVM) outperforms all the other classifiers using unigram features achieving an accuracy rate up to 79.68%. The ensemble classifier successfully identify the overall polarity of the sentence. For example, the polarity of the sentence “I really liked comedy movies” (من فيلم كمدي دوست دَارم) was correctly detect as positive.

4 Conclusion

Sentiment analysis is used for various range of real-world applications such as product reviews, movie reviews, political discussion, etc. However, most of the current research is devoted to English language only, while there are lots of important information available in different languages. In this paper, we propose an ensemble classification using machine learning and deep learning classifiers for Persian sentiment analysis. Experimental results showed that the ensemble classifier achieved better accuracy as compared with deep learning and traditional classifiers. In the future, a more comprehensive and detailed analysis of the proposed ensemble approach (including statistical considerations of each competing system) will be carried out. In addition, we intend to build a novel approach to detect polarity in multilingual sentiment analysis using ensemble classifier.

References

1. Al-Smadi, M., Qawasmeh, O., Al-Ayyoub, M., Jararweh, Y., Gupta, B.: Deep recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels reviews. J. Comput. Sci. **27**, 386–393 (2018)
2. Al-Smadi, M., Talafha, B., Al-Ayyoub, M., Jararweh, Y.: Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews. Int. J. Mach. Learn. Cybern. 1–13 (2018)
3. Alimardani, S., Aghaie, A.: Opinion mining in Persian language using supervised algorithms (2015)
4. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. arXiv:1607.04606 (2016)
5. Cambria, E., Poria, S., Hazarika, D., Kwok, K.: SenticNet 5: discovering conceptual primitives for sentiment analysis by means of context embeddings. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018)
6. Chen, H., Li, S., Wu, P., Yi, N., Li, S., Huang, X.: Fine-grained sentiment analysis of Chinese reviews using LSTM network. J. Eng. Sci. Technol. Rev. **11**(1) (2018)
7. Dashtipour, K., Gogate, M., Adeel, A., Hussain, A., Alqarafi, A., Durrani, T.: A comparative study of Persian sentiment analysis based on different feature combinations. In: International Conference in Communications, Signal Processing, and Systems, pp. 2288–2294. Springer (2017)
8. Dashtipour, K., Gogate, M., Adeel, A., Ieracitano, C., Larijani, H., Hussain, A.: Exploiting deep learning for Persian sentiment analysis. In: International Conference on Brain Inspired Cognitive Systems, pp. 597–604. Springer (2018)
9. Dashtipour, K., Hussain, A., Zhou, Q., Gelbukh, A., Hawalah, A.Y., Cambria, E.: PerSent: a freely available Persian sentiment lexicon. In: International Conference on Brain Inspired Cognitive Systems, pp. 310–320. Springer (2016)
10. Dashtipour, K., Poria, S., Hussain, A., Cambria, E., Hawalah, A.Y., Gelbukh, A., Zhou, Q.: Multilingual sentiment analysis: state of the art and independent comparison of techniques. Cogn. Comput. **8**(4), 757–771 (2016)
11. Dos Santos, C., Gatti, M.: Deep convolutional neural networks for sentiment analysis of short texts. In: Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pp. 69–78 (2014)

12. Dragoni, M., Petrucci, G.: A fuzzy-based strategy for multi-domain sentiment analysis. Int. J. Approx. Reason. **93**, 59–73 (2018)
13. García-Pablos, A., Cuadros, M., Rigau, G.: W2VLDA: almost unsupervised system for aspect based sentiment analysis. Expert Syst. Appl. **91**, 127–137 (2018)
14. Gasparini, S., Campolo, M., Ieracitano, C., Mammone, N., Ferlazzo, E., Sueri, C., Tripodi, G., Aguglia, U., Morabito, F.: Information theoretic-based interpretation of a deep neural network approach in diagnosing psychogenic non-epileptic seizures. Entropy **20**(2), 43 (2018)
15. Gogate, M., Adeel, A., Hussain, A.: Deep learning driven multimodal fusion for automated deception detection. In: 2017 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1–6. IEEE (2017)
16. Gogate, M., Adeel, A., Marxer, R., Barker, J., Hussain, A.: DNN driven speaker independent audio-visual mask estimation for speech separation. arXiv:1808.00060 (2018)
17. Hazarika, D., Poria, S., Gorantla, S., Cambria, E., Zimmermann, R., Mihalcea, R.: Cascade: contextual sarcasm detection in online discussion forums. arXiv:1805.06413 (2018)
18. Ieracitano, C., Adeel, A., Gogate, M., Dashtipour, K., Morabito, F.C., Larijani, H., Raza, A., Hussain, A.: Statistical analysis driven optimized deep learning system for intrusion detection. In: International Conference on Brain Inspired Cognitive Systems, pp. 759–769. Springer (2018)
19. Ieracitano, C., Adeel, A., Morabito, F.C., Hussain, A.: A novel statistical analysis and autoencoder driven intelligent intrusion detection approach. Neurocomputing (2019)
20. Ieracitano, C., Mammone, N., Bramanti, A., Hussain, A., Morabito, F.C.: A convolutional neural network approach for classification of dementia stages based on 2D-spectral representation of EEG recordings. Neurocomputing **323**, 96–107 (2019)
21. Ieracitano, C., Mammone, N., Hussain, A., Morabito, F.C.: A novel multi-modal machine learning based approach for automatic classification of EEG recordings in dementia. Neural Netw. (2019)
22. Kirilenko, A.P., Stepchenkova, S.O., Kim, H., Li, X.: Automated sentiment analysis in tourism: comparison of approaches. J. Travel Res. **57**(8), 1012–1025 (2018)
23. Li, B., Dimitriadis, D., Stolcke, A.: Acoustic and lexical sentiment analysis for customer service calls. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5876–5880. IEEE (2019)
24. Minaee, S., Azimi, E., Abdolrashidi, A.: Deep-sentiment: sentiment analysis using ensemble of CNN and Bi-LSTM models. arXiv:1904.04206 (2019)
25. Nakayama, M., Wan, Y.: Is culture of origin associated with more expressions? An analysis of yelp reviews on japanese restaurants. Tour. Manag. **66**, 329–338 (2018)
26. Onishi, A., Natsume, K.: Overlapped partitioning for ensemble classifiers of P300-based brain-computer interfaces. PloS One **9**(4), e93045 (2014)
27. Öztürk, N., Ayvaz, S.: Sentiment analysis on Twitter: a text mining approach to the syrian refugee crisis. Telemat. Inform. **35**(1), 136–147 (2018)
28. Peng, H., Ma, Y., Li, Y., Cambria, E.: Learning multi-grained aspect target sequence for Chinese sentiment analysis. Knowl. Based Syst. **148**, 167–176 (2018)
29. Poria, S., Hussain, A., Cambria, E.: Concept extraction from natural text for concept level text analysis. In: Multimodal Sentiment Analysis, pp. 79–84. Springer (2018)
30. Poria, S., Hussain, A., Cambria, E.: Sentic patterns: sentiment data flow analysis by means of dynamic linguistic patterns. In: Multimodal Sentiment Analysis, pp. 117–151. Springer (2018)
31. Rogers, A., Romanov, A., Rumshisky, A., Volkova, S., Gronas, M., Gribov, A.: RuSentiment: an enriched sentiment analysis dataset for social media in Russian. In: Proceedings of the 27th International Conference on Computational Linguistics, pp. 755–763 (2018)
32. Shuai, Q., Huang, Y., Jin, L., Pang, L.: Sentiment analysis on Chinese hotel reviews with Doc2Vec and classifiers. In: 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), pp. 1171–1174. IEEE (2018)
33. Sohangir, S., Wang, D., Pomeranets, A., Khoshgoftaar, T.M.: Big data: deep learning for financial sentiment analysis. J. Big Data **5**(1), 3 (2018)