Multi-Emotion Extraction from Text Using Deep Learning

Soghra Lazemi*, Hossein Ebrahimpour-Komleh
Department of Computer Engineering
The University of Kashan, Kashan, Iran
soghra.lazemi@gmail.com, ebrahimpour@kashanu.ac.ir

Received: 2017/12/13 Revised: 2018/02/05 Accepted: 2018/03/16

***Abstract*— Emotions are a part of everyday communications of people and one of the important elements of human nature. We can distinguish a person's emotions from some outcome behaviors such as speech, facial expression, body movements, and gestures. Another outcome behavior is his/her grammar and written method that reflects the inner states of the person. Since people are nowadays more likely to use textual tools to make the connection, emotion extraction from the text has attracted much attention. The majority of methods in this regard consider emotion extraction from the text as a classification problem. Therefore, most studies depend on a huge number of handcrafted features and are done on feature engineering to enhance the classification performance. Considering that a text may include more than one emotion that only one of them is text dominant emotion, we model the emotion extraction problem as a multi-label classification problem by removing the fixed boundaries of emotions. Next, we recognize all the existing emotions in the sentence and in dominant emotion. Our goal is to achieve a better performance only with minimal feature engineering. To this end, we propose a hybrid deep learning model that benefits both CNN and RNN deep models. The experiments are done on a multi-label dataset including 629 sentences with eight emotional categories. Based on the results, our proposed method shows a better performance (about 0.12%) compared with available multi-label learning methods (e.g., BR, RAKEL, and MLkNN).**

Keywords*— Emotion Extraction, Multi-Label Classification, Machine Learning, structural and semantic information, Deep Learning, Natural Language Processing

1. INTRODUCTION

Emotion is a part of human life that has an important role in daily life and effects on decision-making. We make our decisions and do our actions and entertainments based on our emotions including being happy, angry, sad, bored, and etc. Emotions play important role in different contexts such as motivation, perception, cognition, coping, creativity, attention, planning, reasoning, learning, and decision making. Accordingly, emotion recognition and analysis have been the subject of intense research in different fields such as psychology, neurology, behavioral sciences, and computer sciences, especially in human-computer interaction. Picard, the founder of Affective Computing, studied the role of emotions in the field of human-computer interaction and provided some results useful in games, robots, E-learning, and etc. [1].

People sometimes present their emotions in their facial expressions, speech, text, and gesture by conscious or subconscious. In the meantime, most of them choose the text, in the form of emails, commenting about products, weblogs as a communication interface [2]. Therefore, text-based emotion recognition is a great interest to researchers. Emotion recognition from text can be of interest to policymakers, economists, market researchers, and social scientists. Emotional status detection that uses the text can have applications in many areas such as suicide prevention, intelligent tutoring systems, user's authentication, online communication to build smart robots, product review, and emotions application development in linguistics computing [2, 3].

The remainder of this paper is organized as follows. Section 2 surveys related works to emotion models and emotion extraction from text. Section 3 presents our proposed methodology in details. Experiments and results are given in Section 4. Finally, Section 5 concludes the paper.

2. RELATED WORKS

2.1. Emotion Models

We should choose emotion model by attaining to the context in the field of emotion detection and analysis. According to psychological researches, there are a number of theories about that how to express emotions, but the important approaches that have used to emotion analysis include: Categorical models and Dimensional models [4].

Categorical model is based on discrete emotional classes. This model believes that the people have a limited set of discrete basic emotions. Ekman base model is one of these categories that include six basic emotions: ANGER, DISGUST, FEAR, HAPPINESS, SADNESS and SURPRISE [5]. Another model includes basic emotions of TENSION, DEPRESSION, ANGER, VIGOUR, FATIGUE and CONFUSION [6]. However, there is no consensus between theories that select the basic emotions.

Dimensional model shows feelings on a continuous scale and multidimensional spaces. In this model, each emotion has a place in this space. One of this model state-of-the-art samples is Rusell's Circumplex model that distributes emotions in a two dimensions gyrate space (valence dimension and arousal dimension) and the center of it indicates neutral valence and medium arousal (see Fig 1). The valence dimension indicates whether or not a pleasant emotion; and arousal dimension indicates whether

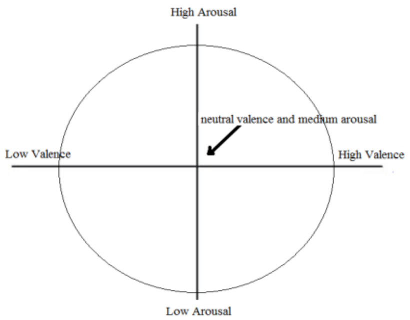

Fig. 1. Graphical Reperesentation of Rusell's Circumplex Model

or not active to do something with this emotion [7]. Another example of this approach is the Mehrabian model that is a three-dimensional model: Pleasure, Arousal, Dominance (PAD). The third dimension indicates whether a person emotion is in control of the situation or not [8].

Categorical model is used because of the simplicity and being more famous. But this model may not cover all the emotions due to limited emotional sets. And the infirmity in this model is the strength of the Dimensional model. Dimensional model is not dependent on a specific emotion and also can consider emotions that are different too low [9].

2.2. *Emotion Detection Methods*

Emotion detection methods divided into three categories: Lexicon-based methods, Rule-based methods and Learning-based methods. Lexicon-based methods are applied at the word level, and they are based on affective lexical resources such as Word Net Affect. Rule-based methods are based on language grammar and structure, and learning-based methods use the properties of the language [10].

Lexicon-based methods are ways that only using one or multiple lexical resources to identify emotions. One of these methods is based on keywords that categorize text to emotional classes using a predetermined set of words. Perikos and Hatzilygeroudis [11] have proposed a simple algorithm that examines emotional words in the sentence and calculates the rate that reflect these emotional words in the text and specified their strength based on the obtained rating. Balahur et al. [12] have presented a method based on ontology that EmotiNet have used as a source to recognize implicit emotions from the context. Statistical methods are also from lexicon-based methods. Gill et al. [13] have been used co-occurrence semantic space techniques such as Latent Semantic Analysis (LSA) and Hyperspace Analogue to Language (HAL) to calculate the semantic similarity between documents and keywords automatically. Rachman et al. [14] have designed a corpus-based approach that the corpus is formed by using the Word Net Affect Emotion and the Affective Norms for English Words. They have used Latent Dirichlet Allocation for expand corpus automatically. In their proposed method, the Ekman emotional model has been used for corpus annotation.

Lexicon-based methods are very simple but rarely used due to the complexity of linguistic structure, such as negative challenge of sentences (a sentence can includes large numbers of negative words, but have a positive meaning). For instance, in the sentence "This book is ridiculously and stupidly beautiful" it can be seen that the repetition of negative words ("ridiculous" and "stupid") is double of positive words ("beautiful"), but the sentence has very positive meaning. However, keywords may also have different meanings and may change based on the context and usage their meaning. Another problem of this method is the ability to understand metaphors and allusions.

In rule-based method, some of rules based on language structure are defined by linguists. Some methods define rules by using a dictionary which contains affective words, (with a list of emotional words by feeling of them) and some without the use of such a dictionary. The disadvantage of this method can be lots of complexity in the design and change the rules, lack of exploring emotions out of defined rules and dependence to language. Shaheen et at. [15] have proposed a method based on syntactic and semantic structure of input sentence. Their proposed method generates rules from input sentences using various ontologies such as Word Net and Concept Net. Then sentences are classified in emotional categories by comparing the generated rules with a set of extracted reference rules from the training set. Anusha and Sandhya [10] have designed a rule-based and learning-based hybrid method. They have defined some rules based on syntactic and semantic features of text. Their proposed method has used natural language processing techniques to improve the performance of learning based classifiers (including SVM and Naïve Bayes). Ekman emotional model and ISEAR database have been used in their research.

Learning-based methods consider the emotion recognition problem as a classification problem and do not need to consider the grammatical structure of the sentence. Of both the supervised and unsupervised machine learning algorithms used to detect emotions. Input data sets must be annotated in supervised methods. It will be done with text annotation by using emotional tags. Annotation is the main supervised method disadvantage in this field. However, the most recent work related to the emotion recognition is in Twitter messages that annotation of these messages is done through hashtags automatically. [16] is one of the first works that models the Categorical model with supervised machine learning algorithms by using SNoW learning architecture. The proposed method has used annotated data sets with Ekman basic emotions. Balabantaray et al. [17] designed emotional classifier using multi-class SVM to recognize Ekman model emotions. Rao et at. [18] have introduced a combinational model based on two supervised intensive topic models. The first model has allocated topics to corresponding emotions and the second model has established association among biterms and emotions by topics. We can refer to [19] that has used the Rusell's Circumplex model and for classifying Twitter messages from the work done using supervised algorithms that have used dimensional model and they have used SVM, KNN algorithms and decision trees.

Unsupervised algorithms are trying to find the hidden structures in unlabeled data to build a model for classifying emotions. The shortcoming of these methods is the lack of structural and semantic information. As well as supervised methods, both categorical and dimensional models also is used in this method. [20] proposed an unsupervised context-based method at the sentence level that does not require any emotional dictionary. Their proposed method has used Ekman basic emotions. In the case of unsupervised methods that has used dimensional model, we can refer to [21] that has used the three-dimensional model; for creating feature vectors they have used Bag of Words (BOW) methods and for reducing the dimensions size they have used LSA, PLSA, NMF methods. Perikos and Hatzilygeroudis [22] have designed Lexicon-based and learning-based hybrid method,. First, the emotional state of a sentence is derived from emotional parts by using a keyword-based approach. The sentence's structures and dependencies are analyzed by using knowledge-based tool. At the end, the emotional state of a sentence is determined by using the statistical machine learning classification methods (including Maximum Entropy and Naïve Bayes).

3. PROPOSED METHOD

Every person's emotion recognition from any of his/her output, including text is a non-fixed behavior, so we can infer different emotions from the text that the text predominant emotion is one of them. The main shortcoming of existing methods is take into account the precise and fixed boundaries for emotions to detect emotions from the text. In other words, existing methods allocate each text only to one of the emotional category. In addition, presence of huge amount of handcrafted features makes emotional detection overwhelming in the existing methods. The aim of this paper is to recognize all the existing emotions in the text and determine the sentence predominant emotion only with minimal feature engineering. To achieve this objective, we present a hybrid deep learning model to learn features automatically. In order to provide sentence level features, our proposed model uses sentence structural information and deep learning models in a hierarchical way. Our proposed method for multi-labeling have the following steps:

1. Sentence segmentation in the separate parts based on conjunction words
2. Word representation by applying word embeddings process
3. Emotion determination using a Convolutional Neural Network (CNN) for each part
4. The sentence's predominant emotion determination using a Long-Short-Term-Memory (LSTM) network

Formally, a sentence consists of n parts as {p1, p2 ,…, pn} and each part includes m words as {x1, x2, …, xm}. According to language grammar, different parts of the sentence usually link each other associated with conjunction words and the used conjunction word is determines the type of connection parts together. So, we divide sentence into parts based on conjunction words in the first phase (if there is any conjunction words). Fig.2 shows an example of this segmentation. Then each part is processed separately and the emotion of each section is extracted using CNN.

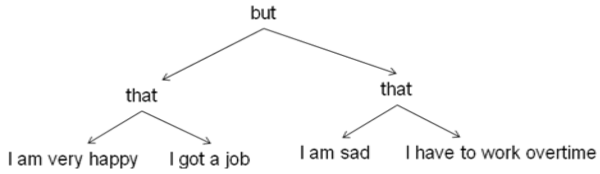

Fig. 2. An Exemple of Sentence Segmentation Based on Conjunction Words

In the second phase, by using word embedding methods we transform words $\{x_1, x_2, \dots, x_n\}$ into continuous vector representations, $\Re^d$, where d is the dimension of the word embeddings. Each part can now be represented as an embedding matrix $M \in \Re^{m\times d}$. We used three convolutional filters to generate dense representations of each part. We set the first, second and third filters $\{K_1 \in \Re^d, K_2 \in \Re^{2d}, K_3 \in \Re^{3d}\}$ to represent unigrams, bigrams and trigrams respectively. Convolution performs on embedding matrix M and the features $c_{j,i}$ is generated by equation (1) [23].

$$c_{j,i} = f\,(x_{i:i+j-1} . K_j^T + b) \tag{1}$$

Where f is the non-linear activation function and b is the bias term.

In order to merge the varying number of features from the convolution layer into a vector, an average pooling layer is used (Equation (2)).

$$C_j = \frac{1}{m}\sum_{i=1}^{m} c_{j,i} \tag{2}$$

Finally, the concatenation of three filters is used as part representation and decoded by a softmax layer into probabilities for each emotion category.

In order to determine predominant emotion of a sentence, LSTM is used in the last phase. LSTM can capture the long-range dependencies in the sentence and let us to use the connections between each part. In other word, we can capture the connection type of each part in the sentence. For a sentence with n parts and $n-1$ conjunction words ($S_i = \{p_1, p_2, \dots, p_n, con_1, con_2, \dots, con_{n-1}\}$), the extracted representation of each part that uses CNN and the existing conjunction word embeddings are fed into a forward LSTM network as input. LSTM calculates the hidden state by taking a combination of three gates: input (i_t), forget (f_t) and output (o_t) gates as follows [24, 25]:

$$\begin{aligned}
x &= \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} \\
f_t &= \sigma(W_f.X + b_f) \\
i_t &= \sigma(W_i.X + b_i) \\
o_t &= \sigma(W_o.X + b_o) \\
c_t &= f_t \otimes c_{t-1} + i_t \otimes \tanh(W_c.X + b_c) \\
h_t &= o_t \otimes \tanh(c_t)
\end{aligned} \tag{3}$$

where ⊗ is element-wise multiplication, W and b select and remove history state vectors and input vectors.

The last output of LSTM includes the previous parts information, so it is used as the final vector. The extracted feature vector is decoded by a linear layer and a softmax layer into probabilities for each emotion category and the category with higher probability is selected as sentence's predominant emotion. Fig.3 details our deep model structure with an example.

4. EXPERIMENTS AND RESULTS

4.1 Data Set

Experiments are done based on the presented database in [26]. This database contains the reviews taken from IMDB and are related to the American History X, The Bourne Identity, Earth (2007), The Godfather, Little Miss Sunshine, The Notebook, SAW and Se7en movies, and each film has at least 6 reviews. The database contains 629 sentences, 8 basic emotions; and 420 sentences include at least one label. The dataset is split into train and test sets using 10-fold cross validation method, 90% of the data is used for training and 10% is used for testing.

4.2 Evaluation Metrics and Results

For comparison, we have used multi-label learning algorithms. In general, multi-label classification algorithms fit in two categories: (1) problem transfer methods (2) algorithm adaptation methods. The first category reduces multi-label classification problem to one or more single classification problem, while the second are trying to adapt existing algorithms for multi-label classification problem. We have compared our proposed method with Binary Relevance (BR), RAndom *k*-labEL (RAKEl) and Multi-label k Nearest Neighbor (MLKNN) algorithms. BR and RAKEL algorithms belong to the first category and MLKNN belongs to the second category. BR transforms the dataset into L separate binary problems (one for each label) and trains with any off-the shelf binary base classifier. RAKEL does LP (Label Powerset) method on M subsets ⊂ {1,2, …, L} of size k. MLKNN has been derived from the traditional KNN algorithm and is the first multi-label lazy learning approach [27].

In our experiments, BR and RAKEL algorithms have been implemented using SVM, as the base classifier. SVM is trained with linear kernel. For RAKEL, k = 3 and M = 2 | L | are considered, because of Tsoumakas et at. [28] Advise. We have considered the number of nearest neighbors equal to 10 since the number of the nearest neighbors used by MLKNN doesn't have a significant impact on the performance of the algorithm. Learning rate=0.001 is used for LSTM. We have used F1-measure to evaluate the performance of system. F-measure shows system performance for each of the labels, so Macro measure is used for overall system performance evaluation (as (Eq.4)):

$$
\begin{aligned}
\mathrm{Pr}\,esicion &= \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i \cap h(x_i)|}{|h(x_i)|} \\
\mathrm{Re}\,call &= \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i \cap h(x_i)|}{|y_i|} \\
F_1 - measure &= \frac{2 * \mathrm{Pr}\,esicion * \mathrm{Re}\,call}{\mathrm{Pr}\,esicion + \mathrm{Re}\,call} \\
Macroaverage\beta &= \frac{1}{q}\sum_{j=1}^{q}\beta(TP_j, FP_j, TN_j, FN_j)
\end{aligned}
\tag{4}
$$

Where L={λ_j, j=1,2,..,q} is the set of all labels, n is the number of examples, Y_i is the ground truth label assignment of the i^{th} examples, x_i is the i^{th} example and $h(x_i)$ is the predicted labels for the i^{th} example.

Each of the sentences is transferred to the vector space in order to use multi-label classification algorithms. The stop words are removed and stems of words is extracted using the Porter stemmer to perform this transferring. We have used the English stop in order to remove stop words.

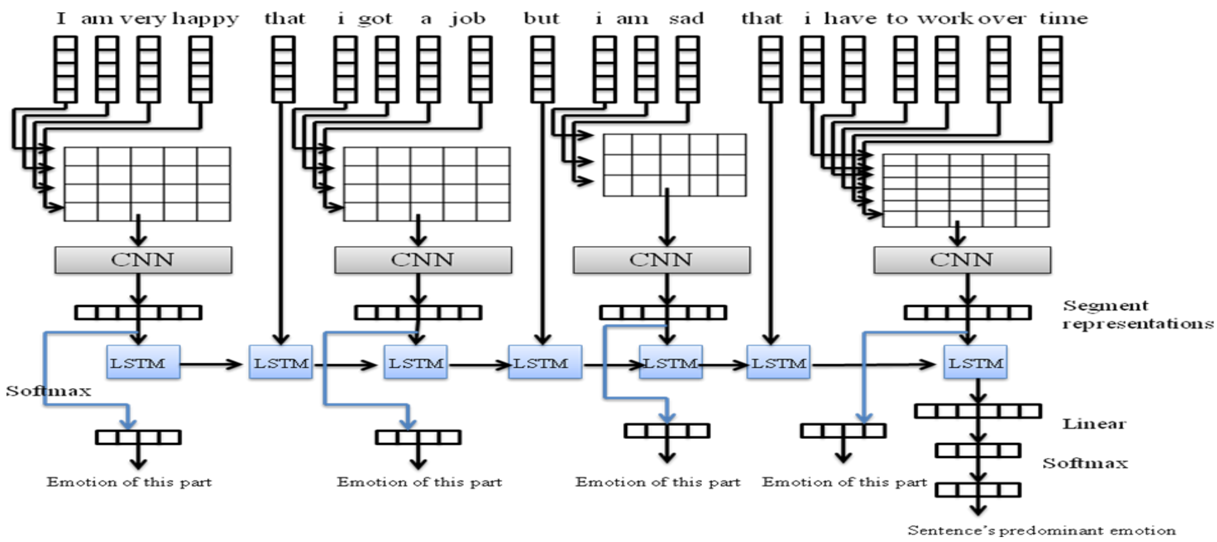

Fig. 3. An Exemple of Multi-Emotion Detection Using Oure Proposed Model

We experimented with two embeddings, namely Glove embeddings and word2vec embeddings. Tables I, II show the prediction performance of the proposed method and multi-label classification algorithms in terms of GloVe-100-dimension and GloVe-300-dimension embeddings respectively. Overall system performance is shown in last row. Our proposed method has the better performance by having a comparison with the results (although not for all categories).

Table III and IV show the obtained results after applying word2vec-100-dimension and word2vec-300-dimension embeddings respectively. As we can see in Tables, in most cases deep learning model has better performance than the multi-label classification algorithms.

Based on Tables I, III and II, IV a large significant improvement not obtained when word2ved embeddings are used. In addition, we captured no significant improvement with the high dimensional embeddings in both cases.

The results indicate that our proposed method has better performance compared to other algorithms. The reason for this is the capture of long-range dependencies information. Multi-label learning methods transfer whole sentence to vector space, so it losses semantic information and connections between different parts of sentence. Our proposed method has maintained sentence's parts semantic connection by using linguistic structural information and has classified each part to proper emotional class by learning method and has reduced the confusion between emotional categories by relations between them that multi-label learning algorithms are faced with them and extracts the correct emotions.

BR and MLKNN algorithms training time is low because of linear correlation with the number of labels and RAKEL training time dependent on the number of made models is slightly longer than the previous models.

5. Conclusion

The main objective of this paper is to extract all existing emotions in a sentence and determine the sentence predominant emotion. We discovered different emotional categories of sentences using a hierarchy of feature extraction in a deep structure neural network. We constructed a hybrid neural network (CNN and LSTM) in a different resolution to extract emotion. The results show that the use of deep learning method can be effective and

TABLE I. PERFORMANCE RESULTS BY GLOVE-100D

Labels	GloVe-100D			
	BR	**RA*K*EL**	**ML*k*NN**	**Our proposed method**
Happiness	0.760	0.711	0.798	0.812
Sadness	0.751	0.726	0.701	0.805
Fear	0.802	0.830	0.796	0.773
Anger	0.672	0.702	0.771	0.814
Disgust	0.769	0.810	0.793	0.890
Surprise	0.518	0.654	0.688	0.900
Macroaverage F_1	0.712	0.738	0.757	0.832

TABLE II. PERFORMANCE RESULTS BY GLOVE-300D

Labels	GloVe-300D			
	BR	**RA*K*EL**	**ML*k*NN**	**Our proposed method**
Happiness	0.731	0.751	0.806	0.832
Sadness	0.767	0.753	0.717	0.809
Fear	0.783	0.830	0.809	0.763
Anger	0.684	0.701	0.775	0.804
Disgust	0.801	0.806	0.802	0.891
Surprise	0.542	0.634	0.689	0.903
Macroaverage F_1	0.718	0.745	0.766	0.833

TABLE III. PERFORMANCE RESULTS BY WORD2VEC-100D

Labels	word2vec-100D			
	BR	**RA*K*EL**	**ML*k*NN**	**Our proposed method**
Happiness	0.762	0.724	0.803	0.814
Sadness	0.749	0.746	0.694	0.813
Fear	0.806	0.802	0.798	0.768
Anger	0.682	0.698	0.725	0.821
Disgust	0.771	0.801	0.815	0.895
Surprise	0.521	0.652	0.704	0.908
Macroaverage F_1	0.715	0.737	0.756	0.836

TABLE IV. PERFORMANCE RESULTS WORD2VEC-300D

Labels	word2vec-300D			
	BR	**RA*K*EL**	**ML*k*NN**	**Our proposed method**
Happiness	0.735	0.740	0.804	0.809
Sadness	0.761	0.743	0.694	0.817
Fear	0.779	0.826	0.799	0.771
Anger	0.691	0.702	0.731	0.834
Disgust	0.812	0.798	0.805	0.887
Surprise	0.541	0.631	0.711	0.919
Macroaverage F_1	0.719	0.740	0.757	0.839

give better results compared with traditional multi-label learning algorithms and does not need a huge number of handcrafted features. Experiment results illustrate that our model superiority overtakes others in term of accuracy. As the future works, we can soften precise and constant borders with fuzzy rules and extract all the emotions in a sentence by fuzzification. Also, semantic consideration using ontology or semantic graph can have more accuracy and better results. According to the collected data sets related to specific areas (film), the high prevalence of some terms and phrases is not unexpected. Finding these features in the text and considering them in multiple tags can increase the accuracy of recognition. Moreover, due to the long-range dependencies existence between each part of the sentence, the review of multi-emotion extraction from text is strongly recommended as a sequence labeling problem.

REFERENCES

[1] R. W. Picard and R. Picard, Affective computing, vol. 252: MIT press Cambridge, 1997.

[2] E. C.-C. Kao, C.-C. Liu, T.-H. Yang, C.-T. Hsieh, and V.-W. Soo, "Towards text-based emotion detection a survey and possible improvements," in Proceedings of International Conference on Information Management and Engineering, pp. 70-74, 2009.

[3] B. Desmet and V. Hoste, "Emotion detection in suicide notes," Expert Systems with Applications, vol. 40, pp. 6351-6358, 2013.

[4] S. Mac Kim, Recognising Emotions and Sentiments in Text, University of Sydney, 2011.

[5] P. Ekman, "An argument for basic emotions," Cognition & Emotion, vo1. 6, pp. 169-200, 2008.

[6] R. Plutchik and H. Kellerman, Emotion, theory, research, and experience: theory, research and experience, Academic press, 1980.

[7] T. Tsujimoto, Y. Takahashi, S. Takeuchi, and Y. Maeda, "RNN with Russell's circumplex model for emotion estimation and emotional gesture generation," in Congress on Evolutionary Computation, pp. 1427-1431, 2016.

[8] A. Mehrabian, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament," Current Psychology, vol. 14, pp. 261-292, 1996.

[9] L. Canales and P. Martínez-Barco, "Emotion detection from text: A survey," in Proceedings of the 5th Conference on Information Systems Research, p. 37-43, 2014.

[10] V. Anusha and B. Sandhya, "A learning based emotion classifier with semantic text processing," Advances in Intelligent Informatics, ed: Springer, pp. 371-382, 2015.

[11] I. Perikos and I. Hatzilygeroudis, "Recognizing emotion presence in natural language sentences," in Proceedings of International Conference on Engineering Applications of Neural Networks, pp. 30-39, 2013.

[12] A. Balahur, J. M. Hermida, and A. Montoyo, "Detecting implicit expressions of sentiment in text based on commonsense knowledge," in Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, pp. 53-60, 2011.

[13] A. Gill, R. M. French, D. Gergle, and J. Oberlander, "Identifying emotional characteristics from short blog texts," in Proceedings of the 30th Conference on Cognitive Science, BC Love, K. McRae, and VM Sloutsky, eds, pp. 2237-2242, 2008.

[14] F. H. Rachman, R. Sarno, and C. Fatichah, "CBE: Corpus-based of emotion for emotion detection in text document," in Proceedings of the 3rd International Conference on Information Technology, Computer, and Electrical Engineering, pp. 331-335, 2016.

[15] S. Shaheen, W. El-Hajj, H. Hajj, and S. Elbassuoni, "Emotion recognition from text based on automatically generated rules," in Proceedings of International Conference on Data Mining, pp. 383-392, 2014.

[16] C. O. Alm, D. Roth, and R. Sproat, "Emotions from text: machine learning for text-based emotion prediction," in Proceedings of Conference on Human Language Technology and Empirical Methods in Natural Language processing, pp. 579-586, 2005.

[17] R. Balabantaray, M. Mohammad, and N. Sharma, "Multi-class twitter emotion classification: A new approach," International Journal of Applied Information Systems, vol. 4, pp. 48-53, 2012.

[18] Y. Rao, J. Pang, H. Xie, A. Liu, T.-L. Wong, Q. Li, et al., "Supervised Intensive Topic Models for Emotion Detection over Short Text," in Proceedings of International Conference on Database Systems for Advanced Applications, pp. 408-422, 2017.

[19] M. Hasan, E. Agu, and E. Rundensteiner, "Using hashtags as labels for supervised learning of emotions in twitter messages," in Proceedings of Workshop on Health Informatics, 2014.

[20] A. Agrawal and A. An, "Unsupervised emotion detection from text using semantic and syntactic relations," in International Conferences on Web Intelligence and Intelligent Agent Technology, pp. 346-353, 2012.

[21] R. A. Calvo and S. Mac Kim, "Emotions in text: dimensional and categorical models," Computational Intelligence, vol. 29, pp. 527-543, 2013.

[22] I. Perikos and I. Hatzilygeroudis, "Recognizing emotions in text using ensemble of classifiers," Engineering Applications of Artificial Intelligence, vol. 51, pp. 191-201, 2016.

[23] Y. Zhang, and B. Wallace, "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification," arXiv preprint arXiv:1510.03820, 2015.

[24] S. Hochreiter, and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, pp.1735-1780, 1997.

[25] A. F. Gers, S. Jürgen, and C. Fred, "Learning to forget: Continual prediction with LSTM," 9th International Conference on Artificial Neural Networks, pp. 850-855,1999.

[26] L. Buitinck, J. Van Amerongen, E. Tan and M. de Rijke, "Multi-emotion detection in user-generated reviews," European Conference on Information Retrieval, pp.43-48, 2015.

[27] M. S. Sorower, "A literature survey on algorithms for multi-label learning," Oregon State University, Corvallis, pp. 1-25, 2010.

[28] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Random k-labelsets for multilabel classification," IEEE Transactions on Knowledge and Data Engineering, vol. 23, pp. 1079-1089, 201

Soghra Lazemi is a PhD candidate of Artificial Intelligence in Kashan University. Her main research areas are text mining, web mining, recommender system and Persian natural language processing.

Hossein Ebrahimpour komleh is an assistant professor of Artificial Intelligence in Kashan University. His work focuses mainly on the text and image processing, computer vision, neural networks and pattern recognition.