Composite Kernels for Public Emotion Recognition from Twitter

Chien-Hung Chen, Yan-Chun Hsing, Yung-Chun Chang

Abstract**—**The Internet has grown into a powerful medium for information dispersion and social interaction that leads to a rapid growth of social media which allows users to easily post their emotions and perspectives regarding certain topics online. Our research aims at using natural language processing and text mining techniques to explore the public emotions expressed on Twitter by analyzing the sentiment behind tweets. In this paper, we propose a composite kernel method that integrates tree kernel with the linear kernel to simultaneously exploit both the tree representation and the distributed emotion keyword representation to analyze the syntactic and content information in tweets. The experiment results demonstrate that our method can effectively detect public emotion of tweets while outperforming the other compared methods.

Keywords**—**Public emotion recognition, natural language processing, composite kernel, sentiment analysis, text mining.

I. Introduction

As social media continues to thrive during the past few years, a spectacular amount of data has been produced online. It has become very valuable and important resources for people to facilitate comprehension public emotions, since people are able to easily disseminate information with emotions through social media [2], [14]. Analyzing public emotions is critical in comprehending the general impression of a given topic. One of the numerous applications of this analysis is the examination of trends in political elections. During the period of an election, a candidate can utilize the public emotions expressed on the social media to capture important issues and make corresponding adjustments in order to gain more support from the general public [15]. For instance, a tool called EMOTIVE [1] exploited Twitter emotion to successfully predict the outcome of the presidential election in the U.S. in 2016. The EMOTIVE system classifies tweets into eight emotions and tracks these emotions throughout a period of time [2]. The research team manipulated the emotion fluctuations toward candidates in EMOTIVE to predict who will be the next president. Therefore, a candidate can take appropriate actions based on the emotions on social media to develop a strong reputation. This case demonstrates that exploring and analyzing social media can be a powerful way of understanding the trends of public emotions.

To recognize emotion behind text is a significant research area in natural language processing (NLP). The research purpose is to analysis opinions that are subjective statements reflecting people’s emotions, perceptions regarding specific topics as well [18]. Al Masum et al. have employed the "sense" emotion from text in news [3], [4], and Quan et al. also used blogs as objects and data sources for Chinese emotional expression analysis [5]. However, most of the contents on social media consist of short texts with about 200 words as microblogging has become a very popular communication tool [6]. Due to the lack of context data, the efficiency of machine learning models is impaired. Therefore, emotion recognition of short text is also increasingly important and challenging in NLP research field [7].

In light of the current social media trend, in this work, we attempt to capture the perception of public emotions on Twitter. We investigated different tree representations of text, and also presented a short text modeling method that utilizes embeddings of emotion keywords to perform public emotion recognition. To detect the public emotions behind tweets, we developed a composite kernel classification method that integrates a smoothed partial tree kernel (SPTK) [11] with a linear kernel to support vector machines (SVM) [19]. The SPTK is developed under a tree structure which is represented by syntactic and content information in the text of tweets, and the linear kernel is developed under our proposed distributed public emotion vector (DPEV). The results of experiments demonstrate that the composite kernel classification method is effective in detecting the emotions of tweets while outperforming many other well-known classification models.

II. Related work

Text is one of the most common ways that people use to convey their feelings, and identifying essential factors within text that affect emotion transition is important for human language understanding. This concept facilitated researches related to emotion recognition and sentiment analysis in the field of natural language process. Many emotion recognition methods (e.g. [12], [13]) were proposed for various applications. For instance, Chang et al. [18] presented a principle-based approach that can learn linguistic patterns from news articles for reader-emotion classification. Since the generated patterns are human-readable, they were further adopted to assist emotional resonance writing. Experiment results show that the proposed method can effectively capture syntactic and context information as well as long-distance

This work was supported by grant MOST106-3114-E-001- 002 and TMU106-AE1-B16 from the Ministry of Science and Technology in Taiwan and Taipei Medical University, respectively.

Chien-Hung Chen and Yan-Chun Hsing are with the Graduate Institute of Data Science, Taipei Medical University, Taiwan (e-mail: m946106007@tmu.edu.tw, m946106006@tmu.edu.tw).

Yung-Chun Chang is with the Graduate Institute of Data Science, Taipei Medical University, Taiwan (corresponding author; e-mail: changyc@tmu.edu.tw).

relation between texts, thereby outperforming other compared methods. Moreover, students indeed improved their writing skills after exploiting the linguistic patterns of reader-emotion. Xiang et al. built a topic-based sentiment mixture model with topic-specific Twitter data integrated in a semi-supervised training framework. The proposed model outperformed the top system in SemEval-2013 [12]. Glorot et al. [13] demonstrated that a Deep Learning system based on Stacked Denoising Auto-Encoders with sparse rectifier units can perform an unsupervised feature extraction which is highly beneficial for the domain adaptation of sentiment classifiers.

Owing to the vigorous growth of social media, researches on social media analytics have gained increasing attention. For example, Xu et al. [14] focused on publicly available Twitter messages through their methods to gather bullying traces. They developed a method to identify the sentiment behind tweets, and further analyzed the topics within tweets with negative sentiment using latent Dirichlet allocation (LDA). In addition, Golbeck et al. [15] predicted user's characteristics based on statistical analysis methods using their collected Facebook dataset. The results show that a user's personality can be accurately predicted through the publicly available information in their Facebook profile.

Our approach differs from the existing emotion recognition methods. To the best of our knowledge, this paper introduces the first composite kernel approach on emotion recognition of short texts. To detect emotion behind tweets effectively, we present each tweet as a lexical centered tree structure and distributed the public emotion vector for the representations of the tree kernel and the liner kernel, respectively. Furthermore, we integrated both kernels to SVMs.

III. A Composite Kernel Approach for Public Emotion Recognition

The system architecture of our method shown in Fig. 1 is comprised of four components, namely *public emotion keyword extraction*, *DPEV representation*, *dependency tree representation*, and *composite kernel classification*. Public emotion recognition is considered as a multi-class classification problem. First, we extract public emotion keywords from a set of tweets. The tweets were then transferred as DPEVs for data representation using the extracted emotion keywords. Meanwhile, each tweet was represented by the dependency tree structure that integrates the syntactic and content information extracted from the tweet. Finally, we adopt the SPTK and the linear kernel for tree representation and distributed keyword vector representation, respectively.

The composite kernel classification component combines both kernels to classify the public emotion of the tweet. Each component is described in detail in the following sub-sections.

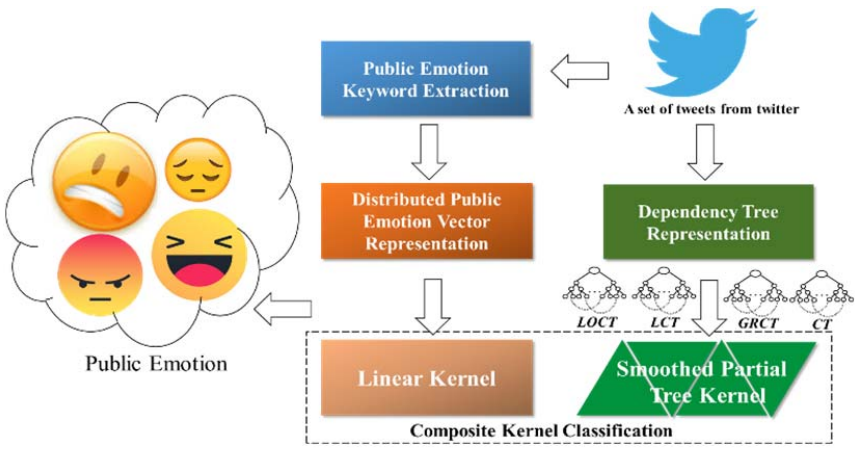

Fig. 1 Systematic architecture of the proposed method for public emotion recognition of tweets

A. Public Emotion Keyword Extraction

Previous text classification studies indicate that using keywords can effectively improve the performance of classification [18], [20]. We implemented the log likelihood ratio (LLR) [21] to capture the keywords in each opinion category. Given a training dataset with emotion categories, we utilize (1) to calculate LLR value of word w in emotion E. The variables of this equation are defined as $k = N(w\wedge E)$, $l = N(w\wedge\neg E)$, $m = N(\neg w\wedge E)$, and $n = N(\neg w\wedge\neg E)$, where $N(w\wedge E)$ denotes the number of documents that contain w and belong to emotion E, $N(w\wedge\neg E)$ denotes the number of documents that contain w but does not belong to emotion E, and so on. We then use (1) to calculate the LLR for w in the emotion E.

$$2\log\left[\frac{p(w|E)^k(1\text{-}p(w|E))^m p(w|\neg E)^l(1\text{-}p(w|\neg E))^n}{p(w)^{k+1}(1\text{-}p(w))^{m+n}}\right] \quad (1)$$

B. DPEV Representation

Keywords extracted for each public emotion category were represented by word embeddings. As shown in Fig. 2, the DPEV representation method is based on combining the emotion keyword embeddings (EKE) from the four emotion categories. Using LLR, we can collect emotion keywords EW for each emotion, where each keyword KW_i is represented by a 300-dimension word embedding. A weight λ_i is assigned to each keyword embedding, and we adopt a weighted average to combine the emotion vectors (EV), which eventually merges the EVs from the angry, fear, sadness, and joy categories. We can thus obtain a 1200-dimension DPEV that can effectively represent the tweet. Take a tweet composed of 10 words (w_1, w_2, …, w_{10}) as an example, where w_5 and w_6 are keywords of angry, and w_8 is the keyword of sadness. The weighted average of keyword embeddings is adopted for representing the angry and the sadness emotions, respectively. The remaining two emotions in which keywords do not exist in this tweet (i.e. fear and joy) will be represented by their average emotion keyword embedding. Finally, we integrate these representations to derive the public emotion keyword embedding, which is a 1200-dimension vector for tweet representation.

NLP and text mining researches often face the problem of data sparseness, especially for the short texts of social media. Therefore, we propose that if there is no emotion keyword in the text, we infer the word embedding through the *K-Nearest Neighbor* model. First, we convert the text into vector via average word embeddings of each word in tweet (with stop words excluded). Next, the five nearest keywords are extracted for representation from the emotion keyword space composed of keywords from the four emotion categories. Since each tweet is represented by the DPEV, we finally use the linear kernel to train a classifier for recognizing the public emotions of tweets.

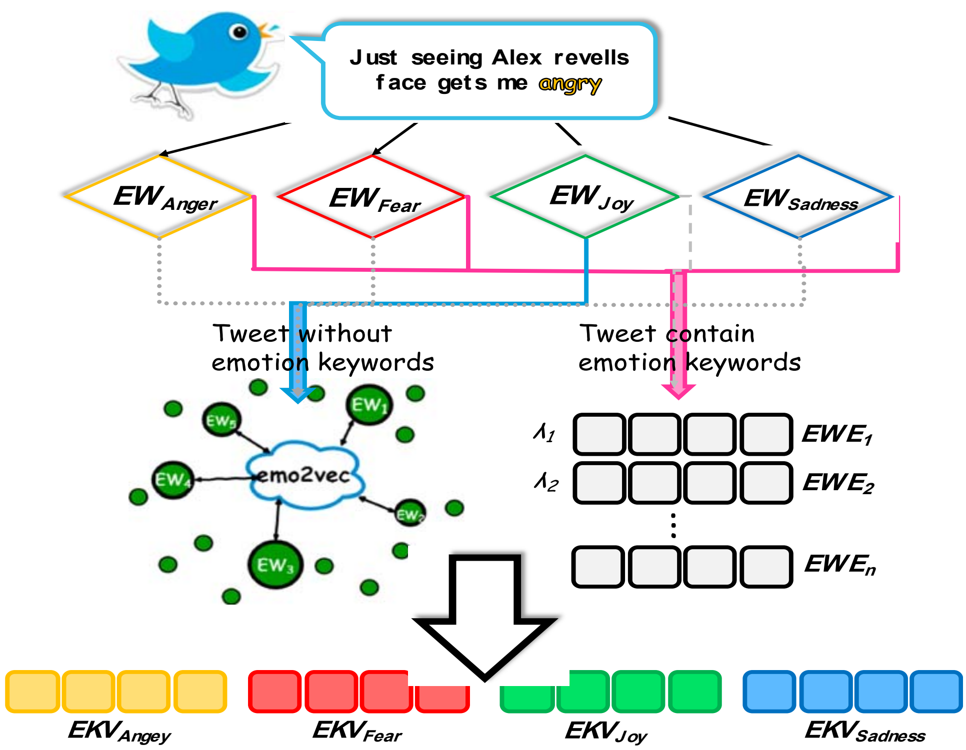

Fig. 2 The DPEV representation

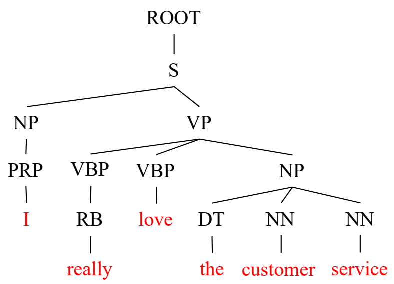

Fig. 3 Constituency tree (CT)

C. Smoothed Partial Tree Kernel

Different tree representations in the tree kernel-based approach may lead to modeling more effective syntactic or semantic feature spaces. In this work, we adopt three kinds of tree structures for representation of tweets, including constituency tree (CT), grammatical relation centered tree (GRCT), and lexical centered tree (LCT). Here, we use a tweet "*I really love the customer service*" as an example for the explanation of three different tree representations. As shown in Fig. 3, a tweet is simply parsed by Stanford Parser for capturing grammatical information called CT. Consequently, Croce et al. [11] proposed GRCT and LCT to complement CT. GRCT and LCT involve grammatical relations (GR), PoS-tags and dependencies. GRCT adds tags of grammatical relations and lexical information as new nodes in CT to emphasize the grammatical relationship information, while LCT enhances the lexical information by adding grammatical relations and PoS-tags as the rightmost children. Figs. 4 and 5 display the GRCT and LCT of the example tweet.

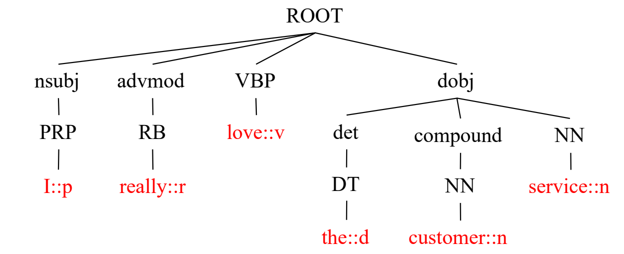

Fig. 4 Grammatical relation centered tree (GRCT)

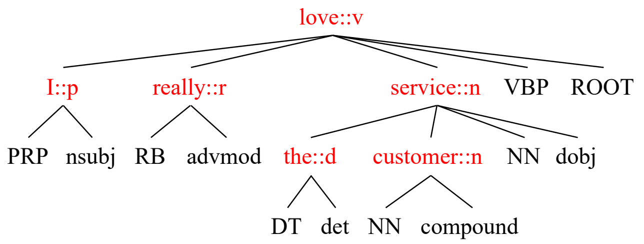

Fig. 5 Lexical centered tree (LCT)

In SVM, a kernel function is employed to cleverly compute the similarity between two instances without requiring the identification of the entire feature space. For tree kernels, it represents tree in terms of their substructures and evaluates the number of common tree fragments between two trees T_1 and T_2 through the following equation:

$$K\left(\mathrm{T}_1, T_2\right) = \sum_{n_1 \in N_{T_1}} \sum_{n_2 \in N_{T_2}} \Delta\left(n_1, n_2\right) \tag{2}$$

where $\Delta(n_1, n_2)$ is the function for calculating the number of common fragments rooted in the node pair n_1 and n_2, which belong to the sets of nodes N_{T_1} and N_{T_2} in T_1 and T_2, respectively. However, it is unable to directly adopt the feature vector since the number of different sub-trees is exponential with the parse tree size. In recent years, multiple tree kernels have been proposed to resolve this computational issue, such as the syntactic tree kernel [8], the partial tree kernel [9], and the lexical semantic kernel [10]. However, the lexical information in these tree kernels must belong to the leaf nodes of the exactly same structures, which restrict their application on dependency trees. Croce et al. [11] proposed a much more general smoothed tree kernel (i.e. SPTK) that can be applied to any tree and exploits any combination of lexical similarities while respecting the syntax enforced by the tree. Therefore, we adopt SPTK to capture the syntactic similarity between the above high dimensional vectors implicitly, as well as the partial lexical similarity of the trees. The $\Delta_{SPTK}(n_1, n_2)$ can be defined as follows:

(1) If nodes n_1 and n_2 are leaves, then $\Delta_{SPTK}(n_1, n_2) = \mu\lambda\sigma(n_1, n_2)$
(2) Otherwise, calculate $\Delta_{SPTK}(n_1, n_2)$ recursively as:

$$\begin{aligned}&\Delta_\sigma\left(n_1, n_2\right) = \\ &\mu\sigma\left(n_1, n_2\right) \times \left(\lambda^2 + \sum_{\vec{I}_1, \vec{I}_2, l(\vec{I}_1) = l(\vec{I}_2)} \lambda^{d(\vec{I}_1) + d(\vec{I}_2)} \times \prod_{j=1}^{l(\vec{I}_1)} \Delta_\sigma\left(c_{n_1}\left(\vec{I}_{1j}\right), c_{n_2}\left(\vec{I}_{2j}\right)\right)\right)\end{aligned} \tag{3}$$

where σ is any similarity between nodes, and μ, $\lambda \in [0,1]$ are two decay factors. The subsequences of child nodes u can be indexed through a set of index sequence $\vec{I} = (i_1, \dots, i_{|u|})$, that is $\vec{I}_1$ and $\vec{I}_2$ represent two sequences of indices in u. In addition, $d(\vec{I}) = i_{|u|} - i_1 + 1$ is the distance between the first and the last child. c is one of the children of the node n also indexed by $\vec{I}$. This provides an advantage in which tree fragments can be matched by applying word embedding similarity σ. As a result, these tree fragments are semantically related even if they are not identical.

D. Kernel Combination

Currently, multiple kernel methods have been widely employed to boost the performance of classification problems [24], [25]. Herein, we implement a composite kernel approach to interpolate the SPTK and the linear kernel to simultaneously exploit different data representations. Polynomial interpolation [26] is used to integrate the two kernels as follows:

$$\begin{aligned}&K_{COM}\left(tw, td\right) = \\ &\alpha \cdot K_{LK}^{P}\left(DEKV_{tw}, DEKV_{td}\right) + (1-\alpha) \cdot K_{SPTK}\left(DT_{tw}, DT_{td}\right)\end{aligned} \tag{4}$$

where *tw* denotes a tweet, *td* is a training instance in the training corpus, and *DEKV* and *DT* are the corresponding DPEV and dependency trees, respectively. $K^P(\bullet,\bullet) = (K(\bullet,\bullet)+1)^d$ is the polynomial expansion of kernel $K(\bullet,\bullet)$. The parameters d and α indicate the polynomial degree and weight coefficient, respectively.

IV. Experiments

A. Experimental Setting

We used the dataset of the WASSA-2017 shared task for emotion intensity (EmoInt-2017) [22] for performance evaluation. There are four different emotional categories including anger, fear, joy, and sadness. The EmoInt dataset is separated into the training, development, and test sets. We merged the training and development sets for training (3960 tweets) and assessed the efficiency of the system using the test dataset (3142 tweets). Distribution of the dataset is shown in Table I.

In our implementation, all tweets were parsed using the Stanford parser [27] to generate the output of parse tree and part-of-speech tagging. In addition, we employed the KeLP package [16] to implement the SPTK classification component and developed three kinds of tree representations. For computing the lexical similarity and constructing the distributed emotion keyword vector, we utilized the pre-trained word embeddings (300-dimension) of Twitter [29]. The evaluation metrics used to determine the relative effectiveness of the compared methods include the precision, recall, F_1-score, and accuracy [17]. We exploit the macro- and micro-averaged score to indicate the overall performance among four different emotional categories for each evaluation metric.

B. Results and Discussion

A comprehensive performance evaluation of the proposed composite kernel approach with other methods is provided. Word embeddings-based approaches which represents each tweet as the average of word embeddings (300-dimension embeddings) and classified by either the SVM [19] (denoted as SVM) or the eXtreme Gradient Boosting model [23] (denoted as *XG*) were included. Next, the standard recurrent neural network [28] method was also developed for evaluation (denoted as *RNN*). In addition, we further compared our method with the SPTK [11] utilizing two different tree representations (denoted as *GRCT* and *LCT*). For the baseline of comparison, the results of Naive Bayes [21] are also included (denoted as *NB*).

TABLE I
DESCRIPTIVE STATISTICS OF THE EMOINT 2017 DATASET

Emotion	# Training	# Test	*Total*
Anger	941 (23.8%)	760 (24.2%)	1701 (23.9%)
Fear	1257 (31.7%)	995 (31.7%)	2252 (31.7%)
Joy	902 (22.8%)	714 (22.7%)	1616 (22.8%)
Sadness	860 (21.7%)	673 (21.4%)	1533 (21.6%)

TABLE II
PERFORMANCE EVALUATION ON PUBLIC MOTION RECOGNITION

System	*Precision, Recall, F_1-score, Accuracy (%)*				
	Anger	Fear	Sadness	Joy	A^W
NB	57.8/60.6/59.2/79.8	57.5/34.9/43.5/71.3	40.4/43.2/41.8/74.2	47.8/68.5/56.3/75.8	51.8/50.5/49.8/75.0
SVM	63.5/17.6/27.6/77.6	36.0/**93.7**/52.0/45.2	63.9/17.1/27.0/80.2	**89.4**/20.0/32.7/81.3	60.7/42.1/36.3/68.7
XG	40.3/33.3/36.5/71.9	40.1/50.2/44.6/60.5	28.2/25.9/27.0/70.1	48.5/44.4/46.3/76.6	39.5/39.6/39.2/69.0
RNN	59.3/67.8/63.2/80.9	61.6/73.1/66.8/77.0	69.5/31.8/43.6/82.4	63.3/69.5/66.2/83.9	63.1/62.2/60.8/80.7
GRCT	41.5/40.4/41.0/71.8	47.7/51.6/49.5/66.7	39.8/40.7/40.2/74.1	48.9/43.7/46.2/76.8	44.8/44.7/44.7/71.8
LCT	61.7/56.8/59.2/81.0	56.5/63.0/59.6/72.9	52.6/55.1/53.8/79.8	66.9/58.8/62.6/84.0	59.3/58.9/58.9/78.9
Our method	**79.3/73.0/76.0/88.9**	**72.3**/79.4/**75.7/83.9**	**77.7/70.4/73.9/89.3**	79.9/**82.8/81.3/91.3**	**76.9/76.7/76.7/87.9**

Table II displays the system performances for public emotion recognition of tweets. As a baseline, the Naïve Bayes classifier is a keyword statistics-based approach which can only accomplish a mediocre performance. In general, each method in this experiment can achieve an overall accuracy around 70%, with the LCT and RNN obtaining better accuracies among the compared methods. Our composite kernel approach can further improve the performance through the combination of multiple kernels, thus achieving the best overall accuracy of 87.9%. In terms of F_1-score, systems simply using the average of word embeddings for data representation can only accomplish an ordinary performance (i.e. SVM and XG). Since these systems concatenated word embeddings of words in a tweet, it is difficult to highlight the discriminative power of words. As a result, the overall F_1-score of both classifiers was below 40%. It is worth noting that the NB classifier indicates keyword information represented by the bag-of-words model, which is crucial in detecting the emotion of tweets and obtained a higher F_1-score than SVM and XG. Moreover, the semantically smoothened tree kernel approaches acquired better performances as they are able to capture the syntax and dependency information of tweets. LCT specifically boosted the overall F1-score to 58.9% since the lexical-centered tree structure can further encode the grammatical relations in the tree structures. The results demonstrate that utilizing lexical features as central node is effective in representing public emotion information in a tweet. In general, the proposed method significantly outperformed all compared systems as it analyzes the semantic (i.e. public emotion keywords), content (i.e. fusion of word embeddings), and syntactic (i.e. lexical-centered tree) information of tweets to identify public emotions. Therefore, our composite kernel method achieved a remarkable performance.

V. CONCLUDING REMARKS

Analyzing public emotions is critical in understanding the general impression of a given topic. It can be achieved through an investigation of social media. In light of this, we extracted the keywords from each emotion category and took advantage of the distributed emotion keyword representation proposed in this work to represent keywords from different perspectives in the vector form. Additionally, the effectiveness of three kinds of dependency tree structures was investigated. We further established a method that combines the dependency tree structure and the linear kernel, which allows the possibility to simultaneously exploit both representations to analyze the syntactic, semantic, and content information in texts. Experiment results demonstrate that the proposed method is effective and also outperformed other well-known classification methods.

In the future, we plan to refine the proposed method and employ it to other NLP applications such as textual entailment and relation extraction. Furthermore, we will investigate the sentimental information in tweets to incorporate more semantic information into both the tree representation and the distributed emotion keyword representation.

REFERENCES

[1] M. D. Sykora, T. W. Jackson, A. O'Brien, and S. Elayan. "Emotive ontology: Extracting fine-grained emotions from terse, informal messages," *International Journal on Computer Science and Information Systems*, vol. 8, no. 2, pp. 106-118, 2013.

[2] O. Miles, et al. "Real-time detection, tracking, and monitoring of automatically discovered events in social media," *In Proceedings of the 52nd ACL*, 2014.

[3] Lin, K. H. Y., Yang, C., & Chen, H. H. "What emotions do news articles trigger in their readers," *In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 733-734, ACM, 2007.

[4] Mostafa Al Masum, S., Prendinger, H., & Ishizuka, M. "Emotion sensitive news agent: An approach towards user centric emotion sensing from the news," *In Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, pp. 614-620, *IEEE Computer Society*, 2007.

[5] Quan, C., & Ren, F. "A blog emotion corpus for emotional expression analysis in Chinese," *International Journal on Computer Speech & Language,* vol. 24, no. 4, pp. 726-749, 2010.

[6] Pak, A., & Paroubek, P. "Twitter as a corpus for sentiment analysis and opinion mining," *In Proceedings of the Language Resources and Evaluation Conference*, vol. 10, 2010.

[7] Rao, Y., Pang, J., Xie, H., Liu, A., Wong, T. L., Li, Q., & Wang, F. L. "Supervised Intensive Topic Models for Emotion Detection over Short Text," *In Proceedings of the Database Systems for Advanced Applications*, pp. 408-422, 2017.

[8] Collins, M., & Duffy, N. "New ranking algorithms for parsing and tagging: Kernels over discrete structures, and the voted perceptron," *In Proceedings of the 40th annual meeting on association for computational linguistics*, pp. 263-270, 2002.

[9] Moschitti, A., "Efficient convolution kernels for dependency and constituent syntactic trees," *In Proceedings of the European Conference on Machine Learning*, pp. 318-329, 2006

[10] Basili, R., Cammisa, M., & Moschitti, A. "Effective use of WordNet semantics via kernel-based learning," *In Proceedings of the Ninth Conference on Computational Natural Language Learning*, pp. 1-8, 2005.

[11] Croce, D., Moschitti, A., & Basili, R. "Structured lexical similarity via convolution kernels on dependency trees," *In Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 1034-1046, 2011.

[12] Xiang, B., & Zhou, L. "Improving twitter sentiment analysis with topic-based mixture modeling and semi-supervised training," *In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, vol. 2, pp. 434-439, 2014.

[13] Glorot, X., Bordes, A., & Bengio, Y. "Domain adaptation for large-scale sentiment classification: A deep learning approach," *In Proceedings of the 28th international conference on machine learning*, pp. 513-520, 2011.

[14] Xu, J. M., Jun, K. S., Zhu, X., & Bellmore, A. "Learning from bullying traces in social media," *In Proceedings of the 2012 conference of the North American chapter of the association for computational linguistics: Human language technologies*, pp. 656-666, 2012.

[15] Golbeck, J., Robles, C., & Turner, K. "Predicting personality with social media," *In Proceedings of the CHI'11 extended abstracts on human factors in computing systems*, pp. 253-262, 2011.

[16] Filice, S., Castellucci, G., Croce, D., & Basili, R. "Kelp: a kernel-based learning platform for natural language processing," *In Proceedings of the ACL-IJCNLP 2015 System Demonstrations*, pp. 19-24, 2015.

[17] Manning, Christopher D., and Hinrich Schütze. Foundations of statistical natural language processing. MIT press, 1999.

[18] Y.-C. Chang, C.-C. Chen, Y.-L. Hsieh, C.-C. Chen, and W.-L. Hsu, "Linguistic template extraction for recognizing reader-emotion and emotional resonance writing assistance," *In Proceedings of the 53rd ACL and the 7th IJCNLP*, pp. 775–780, 2015.

[19] Joachims, T. "Text categorization with support vector machine: Learning with many relevant features," *In Proceedings of the 10th European Conference on Machine Learning*, pp. 137–142, 2005.

[20] Chang, Y. C., Chen, C. C., & Hsu, W. L. "SPIRIT: A tree kernel-based method for topic person interaction detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 9, pp. 2494-2507, 2016.

[21] Manning, C. D., & Raghavan, P. H. Schu tze. "Introduction to Information Retrieval," 2008.

[22] Duppada, V., & Hiray, S. "Seernet at emoint-2017: Tweet emotion intensity estimator," *arXiv preprint arXiv:1708.06185*, 2017.

[23] Chen, T., & Guestrin, C. "Xgboost: A scalable tree boosting system," *In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785-794, 2016.

[24] Miwa, M., Sætre, R., Miyao, Y., & Tsujii, J. I. "Protein–protein interaction extraction by leveraging multiple kernels and parsers," *International journal of medical informatics*, vol. 78, no. 12, pp. 39-46, 2009.

[25] Chang, Y. C., Chen, C. C., & Hsu, W. L. "A composite kernel approach for detecting interactive segments in chinese topic documents," *In Proceedings of the Asia Information Retrieval Societies*, pp. 215-226, 2013.

[26] Zhang, M., Zhang, J., Su, J., & Zhou, G. "A composite kernel to extract relations between entities with both flat and structured features," *In Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pp. 825-832, 2006.

[27] Klein, D., & Manning, C. D. "Accurate unlexicalized parsing,"*In Proceedings of the 41st annual meeting of the association for computational linguistics*, 2003.

[28] Lai, S., Xu, L., Liu, K., & Zhao, J. "Recurrent Convolutional Neural Networks for Text Classification," *In Proceedings of the AAAI*, pp. 2267-2273, 2015.

[29] Pennington, J., Socher, R., & Manning, C. D., "GloVe: Global Vectors for Word Representation," https://nlp.stanford.edu/projects/glove/, 2015.