2020 IEEE International Conference on Big Data and Smart Computing (BigComp)

Emotion Recognition from Text Stories Using an Emotion Embedding model

Seo-Hui Park
Department of Applied Data Science
Sungkyunkwan University
Suwon, Korea
huidea95@skku.edu

Byung-Chull Bae
School of Games
Hongik University
Sejong, Korea
byuc@hongik.ac.kr

Yun-Gyung Cheong
Department of AI
Sungkyunkwan University
Suwon, Korea
aimecca@skku.edu

***Abstract*—In this paper, we analyze emotions in a story text using an emotion embedding model. Firstly, we collected 144,701 tweets, and each tweet is given an emotional hashtag. Using the emotion hashtag as an emotion label, we built a CNN model for emotion classification. We then extracted the embedding model created during the learning process. We then extracted word embedding layer created during the emotion classification learning process.**

We defined this as an 'Emotion embedding model', and applied it to classify story text emotions. The story text used in emotion analysis was ROC story data, and those story sentences are classified into eight emotions based on plutchik's emotion model.

***Index Terms*—Emotion in Text Story, Emotion Analysis in Text, Emotion Embedding Model**

I. INTRODUCTION

Reading emotions in texts written by humans requires high-level intelligence. Studies of textual emotion detection are currently active both in the natural language processing (NLP) and in machine learning (ML) areas. It is because emotions in a story context, along with specific information in the text, play an important role in understanding the given text. Research on emotion analysis is expanding into various applications - e.g., from sentiment analysis of review data to emotional interactive chatbots development [1].

In general, a story consists of three elements - character, event, and setting (i.e., location and temporal setting) [2]. These three factors are crucial information in the field of story text research - including text story generation, text story understanding, and box office predictions from textual film information [3], [4]. Among the three story elements, emotions are closely associated with characters and a sequence of events (i.e., plots), which are essential to story analysis.

In this paper we propose an emotion embedding-based learning model and discuss the results of our experiments in extracting emotional words from a text story dataset and detecting emotions using the proposed emotion embedding model. Emotion embedding model here refers to an embedded layer trained in CNN emotional classification learning process.

II. RELATED WORKS

Studies of textual emotion detection have developed from coarse-grained level to fine-grained level [5]. The Coarse-grained level analysis refers to sentiment analysis which classifies emotions into two polarities - positive or negative. On the other hand, fine-grained level analysis refers to dividing emotions into many emotional states - e.g., Happy, Sadness, Anger and Fear. In this paper we analyze the emotions in a text story on a fine-grained level using eight emotion types based on Plutchik's wheel of emotion model - Anger, Anticipation, Disgust, Fear, Joy, Trust, Sadness, and Surprise [6]. Figure 1 shows 24 emotion types in Plutchik's Wheel of Emotions model, where each basic emotion has 3 levels of intensity (e.g., Terror is more intense than Fear; Apprehension is less intense than Fear). In this paper we focus only on the eight basic emotion types, not taking the intensity of emotions into account.

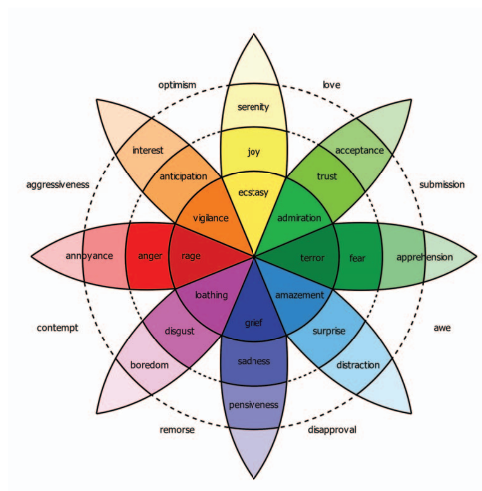

Fig. 1. Plutchik's Wheel of Emotions Model [6]

There are various approaches in analyzing emotions in a text: They can be roughly divided into two categories - a keyword-based method and a learning-based method [7].

A. Keyword-based method

Keyword-based method is the most traditional and easy to use in textual emotional analysis. It first identifies emotional words within a text, and then detects the emotion of a sentence using pre-defined rules and vocabularies. This method utilizes

2375-9356/20/$31.00 ©2020 IEEE
DOI 10.1109/BigComp48618.2020.00014

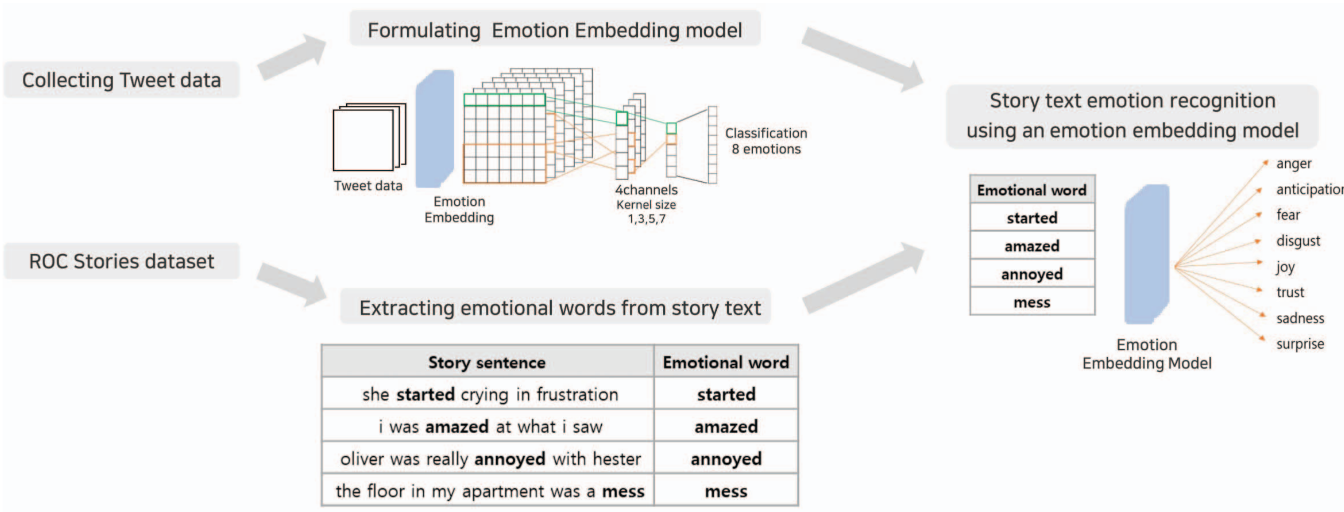

Fig. 2. Emotion Embedding Approach to Textual Story Emotion Classification

rule-based dictionaries with a large number of words and emotional information in them. As a way of recognizing emotions in sentences, the keyword-based method employs emotional scores of each word. Affin [8] vocabulary dictionary, Sentiwordnet [9], and NLTK VADER Sentiment analyzer [10] are typical examples of this method.

B. *Learning-based method*

Learning-based method refers to building a model trained with large amounts of data and then classifying the emotions of test data based on the trained model. Traditional machine learning algorithms such as Naive Bayes, Support vector machines, and decision trees were commonly applied in emotion classification tasks. Furthermore, deep learning algorithms such as Text CNN [11] and LSTM [12] have been applied for emotion recognition within a text. This approach requires sufficient data to be used for learning, and emotional annotations. High-performance emotional classifiers have emerged as massive amounts of data containing emotional annotations have been obtained, such as ISEAR [13], SentiStrengh Twitter (SS-Twitter) [14] and SentiStrengh Youtube (SS-Youtube) [14].

Emotion embedding model is a new technique in the text emotion classification area. Word Embedding is an algorithm that replaces text with vector values, and the types of embedding are divided according to how they are projected into vector spaces. Emotional embedding model is a method that reflects not only co-occurrence and context of words, but also the emotional information of the words when the word is expressed in a low-dimensional vector space.

Seeditabari [15] retrained previous word embedding models (Word2vec [16], GloVe [17], and fasText [18]) to design an emotion embedding model, whereby words with similar emotions are placed closer; and words with opposite emotions are placed far apart. NRC data [19]) containing words and their corresponding emotions are used for this retraining process. Park [20] classified emotions of tweet data using the CNN algorithm, where the embedding model trained by CNN is defined as emotional word vectors (EVEC), and showed that the proposed emotion embedding model performed better in some sentiment classification tasks.

III. Our approach

This section describes our approach to textual emotion analysis by building an emotion embedding model. First, we collect emotional Tweet data with emotion hashtags based on the 8 basic emotion types. Second, using the emotional Tweet data, we build an emotion embedding model. Third, from a story dataset, the most emotional word in each story sentence is extracted using NLTK VADER sentiment analyzer. Finally, based on the extracted word, we detect emotions using our emotion embedding model which are trained with Tweet emotion data. Figure 2 shows the overall procedure for classifying the emotions of a story text with an emotion embedding model.

A. *Collecting Tweet data*

To build an emotion embedding model, a large text datasest with annotated emotions are needed. We employed 144,701Tweet data, where each tweet was annotated with an emotion hashtag. For example, a Tweet, “I broke up with my boyfriend. #sad” in the training data, “ I broke up with my boyfriend” is the declarative sentence part, while #sad is an emotion annotation. Based on Plutchik’s 8 basic emotion types (Anger, Anticipation, Disgust, Fear, Joy, Trust, Sadness, Surprise), we collected Tweets with emotional hashtags consisting of those 8 emotion words or relevant similar words. Table I shows the count of each Tweet data corresponding to 8 basic emotion types.

B. *Building an Emotion Embedding Model*

We conduct a text emotion classification of Tweet data using the CNN learning algorithm. It is a supervised learning

TABLE I
NUMBER OF TWEET DATA EMOTIONS

Tweetdata	
Emotion	Count
Anger	44488
Anticipation	8089
Disgust	8678
Fear	20012
Joy	22489
Trust	10697
Sadness	20462
Surprise	9786
Total	144701

that trains the Tweet data corresponding to 8 emotions and validates the model performance by classifying emotions of the test data. In this learning process, an embedding layer to vectorize text is required. We employ GloVe as the initial embedding model. Then, through the backpropagation process, values of neural network layers are adjusted for emotion classification. The embedding layer is extracted when the best classification performance is reached. The extracted embedding layer has different values compared to the initial embedding layer. In other words, this layer is optimized to detect and classify emotions. We define this layer as our Emotion Embedding layer and apply this model to the emotion detection of text stories.

C. Extracting emotional words from text stories

For text stories, we employ the ROCStories dataset, which include 52,666 stories, each story consisting of 5 simple sentences (263,330 sentences in total) [21]. NLTK VADER Sentiment Analyzer is applied for detecting emotional words in the story sentences. The VADER sentiment analyzer returns positive, negative, neutral and compound scores. The compound score value is an aggregated score of the other three scores- the sum of all the lexicon ratings which have been normalized between -1 and +1: -1 means the highest negativeness; +1 means the highest positiveness. Based on the sentiment polarities of words in the sentences, we select a word with the highest absolute value as a representative emotional word in the given sentence. For example, in a sentence, "one day a guest made him very angry", the polarity of each word is obtained by NLTK VADER Sentiment Analyzer as ([0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.121, 0.5106]). Here, the word "angry" is selected as the emotional word of the sentence with the highest polarity (See Figure 3).

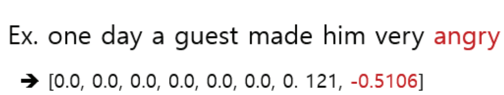

Fig. 3. Example of Selecting a Representative Emotional Word with Highest Sentiment Polarity

D. Textual Story Emotion Recognition Using an Emotion Embedding Model

As the last phase of our pipeline approach, we extract the emotions from text stories using our emotion embedding model. To determine the emotion of each sentence in the stories, we compute the cosine similarity between the selected emotional words using the NLTK VADER sentiment analyzer and emotional hashtags for emotion annotation in Tweet data.

For example, in the story sentence "she was having trouble dealing with her grief", the emotional word extracted from the analyzer is 'grief'. Based on the cosine similarity values, similar words to 'grief' in the emotion embedding model were ['grief', 'sadness', 'sorrow' and despair']. Thus, a story sentence that includes the word 'grief' is classified as the 'Sadness' emotion type.

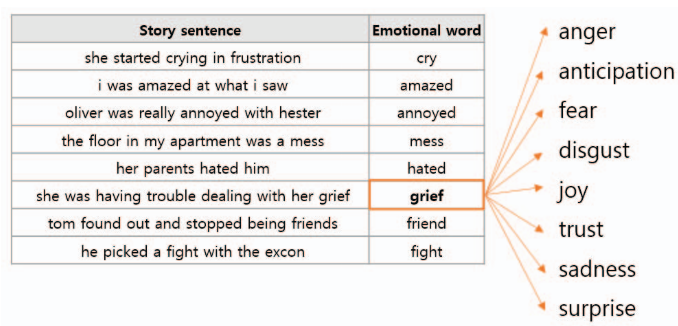

Fig. 4. Computing Similarity Between Emotional words in Text Stories And Emotion Hashtags in Tweet Data

IV. RESULT

This section reports the evaluation results. As shown in Table II, a total of 137,052 story text sentences were analyzed. The classification results show that Joy(22%), Sadness(19.88%), Fear(16.4%) and Anger(14.4%) are the top 4 emotions, occupying a majority of 73.55% of the total counts. These 4 emotions showed higher proportion because the Tweet data used in creating the embedding model also had a lot of Joy, Sadness, Fear, and Anger data (74.26%).

TABLE II
EMOTION ANALYSIS RESULT OF STORY TEXT

Result		
Emotion	Count	Percent
Anger	19711	14.4
Anticipation	13024	9.5
Disgust	3251	2.37
Fear	22480	16.4
Joy	31371	22.89
Trust	6769	4.94
Sadness	27245	19.88
Surprise	13201	9.63

Among the well-classified examples, the sentence "I was amazed at what i saw" was classified as the emotion of

Surprise, because the word ‘amazed’ was extracted as an emotional word. And, in the sentence “Oliver was really annoyed with Hester”, the word ‘annoyed’ was selected as an emotional word, so the emotion of the sentence was classified as Anger, which included ‘annoyed’.

On the other hand, a manual inspection of the results also reveals some mis-classified examples. The sentence “He was determined to play but knew he had to study at night” was classified as the emotion of Joy: the word ‘play’ was selected as the emotional word of the sentence, but the context showed that the emotion of the sentence was not Joy. Likewise, in case of a sentence containing negativity such as ‘not happy’, ‘happy’ was extracted as the emotional word: the sentence was classified as Joy, even though it is the opposite of the actual emotion of the sentence.

story sentence	emotional word	emotion class
she started crying in frustration	cry	sadness
i was amazed at what i saw	amazed	surprise
oliver was really annoyed with hester	annoyed	anger
the floor in my apartment was a mess	mess	disgust
her parents hated him	hated	disgust
she was having trouble dealing with her grief	grief	sadness
tom found out and stopped being friends	friend	sadness
he picked a fight with the excon	fight	anger
she felt prepared	prepared	trust
he decided to play but knew he had to study at night	**play**	**joy**
he was not happy about having to go to school	**happy**	**joy**

Fig. 5. Calculating Simialrity Between Story Emotional Words and Emotion Hashtags

TABLE III
HUMAN ANALYSIS RESULTS - ACCURACY AND FLEISS’ KAPPA SCORE

Emotion	Accuracy	Fleiss’ kappa
Anger	0.367	0.216
Anticipation	0.567	0.363
Disgust	0.55	0.342
Fear	0.483	0.282
Joy	0.733	0.551
Trust	0.517	0.352
Sadness	0.45	0.281
Surprise	0.433	0.293

To evaluate the emotional embedding performance, 120 sentences (15 sentences for each of the eight emotions) of the story sentences were randomly selected. Four human raters evaluated the emotions of the sample sentences. All four raters are college students in their 20s: two males and two females. The raters determined the sentiment polarity of the given sentences and the emotion it belongs to. The Fleiss’ kappa [22] was applied to measure the agreement among the human raters. The result value of is Fleiss’ kappa score returned as a value between 0 and 1 where 0 indicates no consistency and 1 indicates the perfect agreement among the 4 raters.

Based on the evaluation of human raters, we calculated the accuracy of each emotion. The Joy emotion showed the highest accuracy (=.733), while Anger results in the lowest accuracy value (=.367) and the lowest Kappa score (=.216). Inspection of some Anger labeled sentences suggests that these sentences often accompanied other negative emotions such as Sadness and Fear. On the other hand, sentences labeled as Joy did not accompany any other positive emotions.

V. CONCLUSION

In this study, we present a method to extract the emotion of a sentence using an emotion embedding model. For this end, we first built an emotion embedding word model using the collected Tweet data annotated with hashtags. Next, we extract the representative emotional word in each sentence of the ROC story data. The representative emotional word is then used to classify the emotion of the sentence leveraging the cosine similarity. We conducted experiments and the results show that our approach is promising.

We analyze emotions from story texts based on the emotional words representing each story sentence. Therefore, our approach does not take into account the contextual information that can span multiple sentences. It also does not handle expressions negating the sentence such as ‘no’, ‘little’, or ‘not’. In future research, we will enhance our emotion embedding model to detect contextual emotional information in the story text.

ACKNOWLEDGMENT

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2017-0-01642) and Development of QA systems for Video Story Understanding to pass the Video Turing Test (No.2017-0-01772) supervised by the IITP(Institute for Information & communications Technology Promotion). This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST) (No. 2019R1A2C1006316), (No. 2017R1A2B4010499). This work was partly supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST) (No. 2019R1A2C1006316) and Institute of Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (No.2019-0-00421, AI Graduate School Support Program). This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2017-0-01642) supervised by the IITP(Institute for Information & communications Technology Promotion)

REFERENCES

[1] A. Seyeditabari, N. Tabari, and W. Zadrozny, “Emotion detection in text: a review,” *arXiv preprint arXiv:1806.00674*, 2018.

[2] S. B. Chatman, *Story and discourse: Narrative structure in fiction and film*. Cornell University Press, 1980.

[3] K. Pichotta and R. Mooney, “Statistical script learning with recurrent neural networks,” in *Proceedings of the Workshop on Uphill Battles in Language Processing: Scaling Early Achievements to Robust Methods*, 2016, pp. 11–16.

[4] J. Eliashberg, S. K. Hui, and Z. J. Zhang, “Green-lighting movie scripts : Revenue forecasting and risk management,” 2010.

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

[6] R. Plutchik, "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," *American scientist*, vol. 89, no. 4, pp. 344–350, 2001.

[7] K. Sailunaz, M. Dhaliwal, J. Rokne, and R. Alhajj, "Emotion detection from text and speech: a survey," *Social Network Analysis and Mining*, vol. 8, no. 1, p. 28, 2018.

[8] F. Å. Nielsen, "Afinn," *Richard Petersens Plads, Building*, vol. 321, 2011.

[9] A. Esuli and F. Sebastiani, "Sentiwordnet: A publicly available lexical resource for opinion mining." in *LREC*, vol. 6. Citeseer, 2006, pp. 417–422.

[10] C. J. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Eighth international AAAI conference on weblogs and social media*, 2014.

[11] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.

[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[13] H. G. Wallbott and K. R. Scherer, "How universal and specific is emotional experience? evidence from 27 countries on five continents," *Information (International Social Science Council)*, vol. 25, no. 4, pp. 763–795, 1986.

[14] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas, "Sentiment strength detection in short informal text," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 12, pp. 2544–2558, 2010.

[15] A. Seyeditabari, N. Tabari, S. Gholizadeh, and W. Zadrozny, "Emotional embeddings: Refining word embeddings to capture emotional content of words," *arXiv preprint arXiv:1906.00112*, 2019.

[16] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in neural information processing systems*, 2013, pp. 3111–3119.

[17] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.

[18] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135–146, 2017.

[19] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word–emotion association lexicon," *Computational Intelligence*, vol. 29, no. 3, pp. 436–465, 2013.

[20] J. H. Park, "Finding good representations of emotions for text classification," *arXiv preprint arXiv:1808.07235*, 2018.

[21] N. Mostafazadeh, N. Chambers, X. He, D. Parikh, D. Batra, L. Vanderwende, P. Kohli, and J. Allen, "A corpus and evaluation framework for deeper understanding of commonsense stories," *arXiv preprint arXiv:1604.01696*, 2016.

[22] M. L. McHugh, "Interrater reliability: the kappa statistic," *Biochemia medica: Biochemia medica*, vol. 22, no. 3, pp. 276–282, 2012.