CARER: Contextualized Affect Representations for Emotion Recognition

Anonymous Submission

1 Introduction

Text-based emotion analysis is a derivative of sentiment analysis where the objective is to extract the fine grained emotions expressed in textual information [18][2][1]. We propose a method that leverages graph theory and modern deep learning methods to build state of the art text-based emotion recognition systems. The proposed system, called CARER, automatically learns and constructs emotion-relevant, graph-based structural representations and enriches them via word embeddings. The output is then fed as input into a multi-layer convolutional neural network (CNN) for extracting abstract affect representations. Our approach, while effective at recognizing affect states in text, aims to address some of the outstanding challenges in the field of emotion recognition such as *interpretability* and *robustness*. Our main contributions are: 1) A graph-based algorithm for extracting highly interpretable, emotion-relevant pattern-based features, 2) a set of emotion-rich representations enhanced through word embeddings, 3) and a comprehensive performance analysis of various conventional learning models and deep learning models used for text-based emotion recognition.

Emotion recognition systems represent an important layer of abstraction in modern artificial intelligence (AI) systems. They can be leveraged to build empathy-aware conversational bots that have tremendous implications in building real world health applications. For instance, woebot [1] is a chatbot that is able to constantly monitor mental health status based on mood patterns extracted from users' conversations (usually patients). Emotion intelligence is also embedded into other commonly used applications such as recommendation systems [4] and dialogue generation [3]. Emotion recognition systems can also be used to understand other human and linguistic behaviors represented in the form of hate speech [6], sarcasm [5], intent, mental disorders [7], among others.

2 Methodology

In this section we briefly discuss the main components of the proposed system named CARER.

Data: We collect three Twitter datasets using hashtags as noisy labels and annotated through *distant supervision*. The hashtags are emotion-relevant terms such as *#joy* and *#sad* that represent one of eight emotions from Plutchik's wheel of emotions [9]. Two of the datasets (2+ million instances each) are used to construct syntactic patterns based on a graph-based approach. The third dataset is used for training the emotion recognition models so as to avoid any bias arising from the first two datasets.

Graph-Based Pattern Extraction: In this step, the goal is to extract contextualized syntactic patterns with emotion-relevant information. We first build two types of weighted directed graphs, named subjective graph and objective graph, and perform a graph aggregation procedure based on pair-wise adjustment, which results in an emotion graph – a graph consisting of arcs and nodes with stronger emotion relevance. After the aggregation phase, two graph analysis techniques, namely eigenvector centrality and clustering coefficient, are performed on the emotion graph, resulting in two types of nodes, respectively: connector words (typically function words such as *"or"*, *"my"*, and *"and"*) and subject words (typically topical words such as *"never"* and *"life"*). The two sets of words are used as the building blocks to extract emotion-relevant syntactic patterns, achieved by combining

[1] https://woebot.io/

	Document	GT	DeepMoji	EmoNet	CARER_{EK}	Enriched Patterns
Short text	damn what a **night**	joy	surprise	sadness	joy	**what a** { *day*, *rush*, *pass* }
	want it to snow	joy	sadness	fear	joy	{ *need*, *hoping* } **it**
Rare words	**whaaaaaat.** i did not expect that at all	surprise	sadness	fear	surprise	{ *wondering*, *what* } **i**
Mixed emotions	got **thee** worst sleep ever	anger	sadness	sadness	anger	**got** { *madd*, *thatt*, *bacc* }
	what the fuck is **going** on !?	fear	anger	sadness	fear	**is** { *ends*, *finishes* }

Figure 1: Classified documents extracted from the testing data. Words in **bold blue** correspond to the subject words of one of the enriched patterns extracted from the document. Words in *italic green* represent other relevant word examples found in the cluster subject words belongs to. Words in **bold pink** denote the connector word/s in the pattern; E.g., *"it"* is the connector word of patterns *"need it"* and *"hoping it"*. **GT** stands for ground truth.

permutations of connector words and subject words and exhaustively searching patterns in a dataset. Using a modified version of tf-idf [10], the top most important patterns are retained.

Pattern Enrichment: The patterns extracted in the previous step can be further enhanced to make them relevant to an emotion classification task. We chose to use a enrichment process that essentially utilizes word embeddings to retain patterns that share similar subject words based on their vector representations. This procedure also helps to reduce the number of patterns extracted from the previous step, which is useful in cases where there is low computing resources. Given the resulting syntactic patterns, a pattern-to-emotion matrix representation (also referred to as **emovec** representation) is extracted for each tweet post in a different training dataset and fed into a CNN for training.

Model and Training: The emovec representations are fed as input to a multi-layer CNN with two 1-D convolutional layers of sizes 3 and 16. ReLU [11] is used as the activation function followed by 1-max pooling layers. Subsequently, two hidden layers of size 512 and 128 are employed, each applied dropout [2] for regularization. We chose a batch size of 128 and trained for 4 epochs using Adam optimizer [13]. A softmax function is used to generate the final emotion classifications.

3 Experimental Results

We compared the results of our method against various emotion recognition systems. **BoW** represents a bag of words approach enhanced with tf-idf weighted features. **ngram** and **char_ngram** use n-gram features consisting of contiguous word sequences and character sequences, respectively. **LIWC** consists of features extracted from the affect dimensions of the LIWC lexicon [14]. $\textbf{CNN}_{\textbf{w2v}}$ uses the proposed CNN model and word embeddings obtained from [15]. As shown in Table 1, our proposed model, **CARER**, outperforms (F1 average score of 79% and 81% on 8 and 6 emotions, respectively) various traditional approaches (**BoW**, **ngram**, **LIWC**, and **char_ngram**), modern methods ($\textbf{CNN}_{\textbf{w2v}}$), and state of the art approaches (**DeepMoji** [17] and **EmoNet**). **DeepMoji** and **EmoNet**, which are both neural-based approaches, mostly suffer from lack of sufficient contextualization which the graph-based extraction component of our approach helps to address. We also experimented with a Chinese dataset and outperformed a model trained on fastText [16] by an F1 average score difference of 8%. As shown in the examples presented in Figure 1, the contextualized patterns also performed robustly under various different settings important in emotion recognition such as concomitancy (i.e., mixed emotions), short text classification, and implicit emotional expressions.

Models	Features	F1 Avg. (8 emotions)	F1 Avg. (6 emotions)
BoW	**TF-IDF**	0.60	-
ngram	**TF-IDF**	0.63	-
char_ngram	**TF-IDF**	0.57	-
LIWC	**affective words**	0.35	-
$\textbf{CNN}_{\textbf{w2v}}$	**word embeddings**	0.65	0.69
EmoNet	**word embeddings**	0.52	0.58
DeepMoji	**word embeddings**	0.59	0.63
CARER	**enriched patterns**	**0.79**	**0.81**

Table 1: Comparison of our model against various emotion recognition systems.

References

[1] Saravia, Elvis, Hsien-Chi Toby Liu, and Yi-Shin Chen. DeepEmo: Learning and Enriching Pattern-Based Emotion Representations. arXiv preprint arXiv:1804.08847, 2018.

[2] Hinton, Geoffrey E., et al. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580, 2012.

[3] Zhou, Xianda, and William Yang Wang. MojiTalk: Generating Emotional Responses at Scale. arXiv preprint arXiv:1711.04090, 2017.

[4] Ayata, Deger, Yusuf Yaslan, and Mustafa E. Kamasak. Emotion Based Music Recommendation System Using Wearable Physiological Sensors. IEEE Transactions on Consumer Electronics, 2018.

[5] Poria, Soujanya, et al. A deeper look into sarcastic tweets using deep convolutional neural networks. arXiv preprint arXiv:1610.08815, 2016.

[6] Watanabe, Hajime, Mondher Bouazizi, and Tomoaki Ohtsuki. Hate Speech on Twitter: A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection. IEEE Access 6 (2018): 13825-13835.

[7] Huang, Yen-Hao, Lin-Hung Wei, and Yi-Shin Chen. Detection of the Prodromal Phase of Bipolar Disorder from Psychological and Phonological Aspects in Social Media. arXiv preprint arXiv:1712.09183, 2017.

[8] Go, Alec, Richa Bhayani, and Lei Huang. Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford 1.12, 2009.

[9] Plutchik, Robert. 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 89.4 (2001): 344-350.

[10] Leskovec, Jure, Anand Rajaraman, and Jeffrey David Ullman. Mining of massive datasets. Cambridge university press, 2014.

[11] Nair, Vinod, and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. Proceedings of the 27th international conference on machine learning (ICML-10), 2010.

[12] Saravia, Elvis, Carlos Argueta, and Yi-Shin Chen. Unsupervised graph-based pattern extraction for multilingual emotion classification. Social Network Analysis and Mining 6.1 (2016): 92.

[13] Kingma, Diederik P., and Jimmy Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

[14] Tausczik, Yla R., and James W. Pennebaker. The psychological meaning of words: LIWC and computerized text analysis methods. Journal of language and social psychology 29.1 (2010): 24-54.

[15] Deriu, Jan, et al. Swisscheese at semeval-2016 task 4: Sentiment classification using an ensemble of convolutional neural networks with distant supervision. Proceedings of the 10th international workshop on semantic evaluation. No. EPFL-CONF-229234, 2016.

[16] Bojanowski, Piotr, et al. Enriching word vectors with subword information. arXiv preprint arXiv:1607.04606, 2016.

[17] Felbo, Bjarke, et al. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. arXiv preprint arXiv:1708.00524, 2017.

[18] Abdul-Mageed, Muhammad, and Lyle Ungar. Emonet: Fine-grained emotion detection with gated recurrent neural networks. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017.