Using Domain Knowledge to Enhance Deep Learning for Emotional Intelligence (Extended Abstract)

Hortense Fong and Vineet Kumar

Yale School of Management, New Haven CT 06511, USA
{hortense.fong,vineet.kumar}@yale.edu

We propose a hierarchical classification architecture for identifying granular emotions in unstructured text data. Whereas most existing emotion classifiers focus on a coarse set of emotions, such as Ekman's six basic emotions (e.g., joy, anger, sadness) [2], we focus on a larger set of 24 granular emotions (e.g., irritation, envy, rage). Compared to coarse emotions, granular emotions are more specific in the information they convey (e.g., intensity, context). For example, sadness is a broad bucket of emotions that encompasses more specific granular emotions ranging from disappointment to neglect to sympathy. Individuals who are able to better recognize the nuance of their emotional state and capture it with more specific words [3] are typically characterized as having greater emotional intelligence [4, 1].

Granular classification is challenging because it is a fine-grained classification problem, which aims to distinguish subordinate-level categories. It is challenging when there is small inter-class variation but large intra-class variation. In the case of emotions, individuals may use different word patterns to evoke the same emotion and similar word patterns to evoke differing emotions. The underlying idea to overcoming this challenge is to divide the data into similar subsets and then train a separate classifier for each subset so that the model can learn to more easily differentiate similar groups. Motivated by psychology literature, we use the idea of hierarchical classification to improve the identification of granular emotions.

The proposed classifier takes advantage of the semantic network of emotions from the seminal work of Shaver et al. (1987), which maps out how individuals categorize emotions [5]. The semantic network contains a level of coarse emotions and a level of granular emotions that are subordinate to the coarse emotions. The coarse level helps us to divide the data into similar subsets of coarse emotions. Building on this, we develop a classifier that first classifies input text into one or more of the coarse emotions, capturing the idea of mixed emotions, and subsequently classifies the input text into a granular emotion within the coarse emotion(s) identified. The first level is a multi-label classifier and the second level is a multi-class classifier.

We collect self-labeled English tweet data in which the author has included an emotion hashtag to train our classifier. The emotion hashtags we use come from the emotion words in each of the granular emotion clusters identified in [5]. Table 1 shares a few example tweets from our data set. We take a number

 In: N. Chhaya, K. Jaidka, J. Healey, L. H. Ungar, A. Sinha (eds.): Proceedings of the 3rd Workshop of Affective Content Analysis, New York, USA, 07-FEB-2020, published at http://ceur-ws.org

of steps to filter out uninformative tweets and to pre-process the tweets for use in classification. In total, our data set contains 867,264 tweets.

Table 1. Example Tweets

Coarse Emotion	Granular Emotion	Tweet
Love	Affection	The way his eyes look right before he goes to kiss me again.. Oh, I love that. #affection #handsome #hazeleyes
Anger	Rage	For the 2nd time @verizon has erased information from a phone on my account. This time EVERYTHING. Rep said she backed it on the cloud but backed nothing!!! @VerizonSupport #furious
Sadness	Shame	Stayed up til midnight last night baking 30 chocolate chip cookies and 30 snowballs for my friends' Christmas presents. I have two tests today #regret
Sadness	Neglect	I could really use some of my friends right about now :(#lonely #upset #sad

We train deep neural network classifiers since they have demonstrated state-of-the-art performance on emotion classification. The inputs to the neural networks are the tweet text and granular emotion labels. We compare the performance of convolutional neural networks (CNNs) with variants of long short-term memory networks (LSTMs) to show the impact of using a hierarchical classifier versus a flat classifier. For the flat classifier, a separate binary classifier is trained for each granular emotion, resulting in 24 classifiers. For the hierarchical classifier, a separate binary classifier is trained for each coarse emotion and then a multiclass classifier is trained for each coarse emotion to differentiate the granular emotions, resulting in 12 classifiers.

Our performance metrics of interest are precision, recall, and F1. In many applications, recall is the measure of interest because false negatives are more costly. For example, if a customer is feeling exasperated but this sentiment goes unnoticed, the firm risks losing the customer. False positives, on the other hand, are typically less costly. If a happy customer gets tagged as irritated, the firm can easily realize the mislabeling and choose not to act on the tag.

We report the results of each architecture in Table 2. The proposed hierarchical classifier outperforms a single-stage flat classifier in terms of F1 by increasing recall at the cost of precision. For example, the F1 for bi-LSTM increases from 34.93% to 39.84%. We believe the overall F1 from the bi-LSTM can be improved with additional fine-tuning or through the exploration of additional hierarchical neural networks (e.g., BERT).

The hierarchical structure increases the interpretability of the model by enabling interpretation at two levels rather than just one. At one level, it can identify which words contribute to or take away from the positive classification

Table 2. Performance of Classifiers (%)

Classifier	Accuracy	Precision	Recall	F1
Flat				
CNN	96.53	60.09	21.84	31.07
LSTM	96.64	61.19	24.80	34.05
Bi-LSTM	96.67	61.69	25.65	34.93
Hierarchical				
CNN	99.10	52.24	29.59	36.50
LSTM	99.26	51.86	31.42	37.60
Bi-LSTM	99.37	52.12	33.99	39.84

of each of the coarse emotions. At the second level, it can identify which words within each coarse emotion category contribute to each of the granular emotions. Not only does opening up the black box help address model validity, but it also provides insight to end users on which specific terms are diagnostic of each granular emotion.

Overall, we find that the use of domain knowledge to inform the design of a granular emotion classifier through a hierarchical structure improves recall and F1 as well as model explainability.

References

1. Brackett, M.A., Rivers, S.E., Bertoli, M.C., Salovey, P.: Emotional intelligence. In: Handbook of emotions, pp. 513–531. Guilford Publications (2016)
2. Ekman, P., Friesen, W.V., Ellsworth, P.: Emotion in the human face: Guidelines for research and an integration of findings (1972)
3. Lindquist, K.A., Gendron, M., Satpute, A.B.: Language and emotion: Putting words into feelings and feelings into words. In: Handbook of emotions, pp. 579–594. Guilford Publications (2016)
4. Salovey, P., Mayer, J.D.: Emotional intelligence. Imagination, cognition and personality **9**(3), 185–211 (1990)
5. Shaver, P., Schwartz, J., Kirson, D., O'connor, C.: Emotion knowledge: further exploration of a prototype approach. Journal of personality and social psychology **52**(6), 1061 (1987)