Sensory Features in Affective Analysis: A Study Based on Neural Network Models

Abstract. This study proposes an ensemble model to incorporate sensory features of lexical items in English from external resources into neural affective analysis frameworks. This allows the models to take the combined effects of bi-directional feeling between the sensory lexicon and the writer to infer human affective knowledge. We evaluate our model on two affective analysis tasks. The ensemble model exhibits the best accuracy and the results with 1% F1-score improvement over the baseline LSTM model in the sentiment analysis task. The performance shows that perceptual information can contribute to the performance of sentiment classification tasks significantly. This study also provides a support for the linguistic finding that correlations exist between sensory features and sentiments in the language.

Keywords: Affective analysis, sensory feature, sentiment, English

1 Introduction

Affective analysis, a broader term for sentiment analysis and emotion recognition, is highly demanded in the social media text analysis. It is also crucial for various applications, such as opinion based product recommendation [3], opinion mining [7], and medical artificial intelligence (AI) [12]. Although affective analysis has been studied extensively using different methods applied on different types of data, text is one of the most important types of data so far [6]. Existing research on affective analysis mainly centers on learning features through the use of neural networks, such as Convolutional Neural Network (CNN) [5], Gated Recurrent Unit (GRU) [17], Long-Short Term Memory Network (LSTM) [2], Memory network [8], and Pre-trained transformer strategies [4]. Those models, though achieving promising results, have also presented drawbacks in terms of explainability, as their grey box approaches are unable to highlight the salient words or phrases that link to the affective information and thus contribute less to reflect human-understandable emotion components [13]. However, the neural networks approach, which learns from large-scale raw data while devaluing a myriad of existing external resources (e.g., linguistic ontology, cognition-grounded data, and affective lexicon), has showed great usefulness in feature engineering approach and could thus complement the automatic learn features for the neural model to human cognition process. Furthermore, with the recent neural cognitive trend in Natural Language Processing (NLP), it has been shown that modern language algorithms partially converge towards brain-like solutions among all NLP tasks, where affective analysis is one of the most cognition driven tasks [1]. Specifically, two phenomena rally behind the cognitive theories of affective analysis [9]. First, people react to the same event with a variety of different emotions, where the reaction is subject to individuals' biases based on their cognitive experiences. Second, different events may trigger the same emotion, as there are only a limited number of emotional reactions cognitively. Thus, cognition grounded data obtained in crowdsourcing should be helpful in building a cognitive driven neural model for the affective analysis.

This study incorporates sensory features of lexical items from external resources into neural affective analysis frameworks. This allows the models to take the joint effects of bi-directional feeling between the external knowledge and the text to infer human affective knowledge. To the best of our knowledge, this study is one of the first to explore the coupled effects of manually annotated affective vectors and neural networks for the affective analysis, which somehow involves the implicit engagement of human readers to help sentiment prediction. In the rest of our paper, Section 2 gives a review of sensory features in language resources, and Section 3 presents our models based on the sensory features. To obtain empirical evidence, Section 4 shows a series of experiments based on five benchmark datasets with comprehensive comparisons on different models. The last section is conclusion.

2 Sensory features in affective analysis

The sensory features employed in this current study are mainly derived from the sensory rating task on lexical items, which has been initiated by Lynott and Connell [10]. In the task, participants were asked to rate the extent to which they can experience something denoted by a lexical item by feeling through touch, by tasting, by seeing, by hearing, and by smelling. Lynott et al. [11] have added one more dimension for the sensory features of English lexicon (i.e., the modality of interoception).

In terms of the correlation between sensory features and sentiments in the language, Winter [14] is one of seminal work. The study has found that sensory features related to taste and smell are more emotion loaded for English lexicon. Similarly, this pattern has also been observed for Mandarin Chinese [18], except that gustatory features are more positive, while olfactory features are more negative. Thus, this study presumes that the correlations between sensory features and sentiments should be incorporated into neural network models for the affective analysis in the text.

3 Our model

This section elaborates how the affective awareness based on the perceptual features in the English lexicon is employed and incorporated into the neural sentiment classification frameworks for the affective analysis.

3.1 External sensorimotor resource

The external resource utilized by this study is The Lancaster Sensorimotor Norms [11], which was collected through the ratings of 3,500 individual participants of English native speakers using Amazon's Mechanical Turk [1] platform. The Lancaster Sensorimotor Norms covers the annotations in six perceptual modalities (touch, hearing, smell, taste, vision, and interoception) and five action effectors (mouth/throat, hand/arm, foot/leg, head excluding mouth/throat, and torso) for 39,707 lexical concepts. The range of the vector is from 0 to 5. As this study focuses on the affective analysis, we only employ the perceptual information in the six modalities, where the value of each dimension is denoted as $v_j, j \in [0, 5]$. For example, the vector of "blue" is [0.25, 0, 0.15, 0.5, 0, 4.45], corresponding to the auditory, gustatory, haptic, interoceptive, olfactory, and visual score.

3.2 Model structure

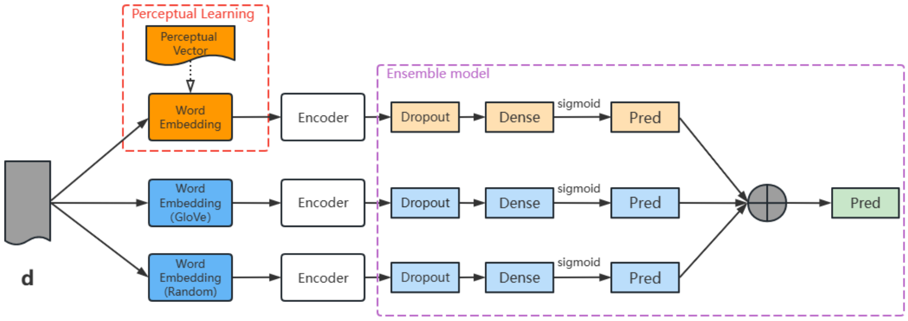

Fig. 1. The architecture of our model

Input and output. The goal of sentiment analysis is to assign an affective label for a piece of input texts, reflecting the writers' attitude. The label types can either be binary for polarity indication or numerical for both polarity and strength. Let **D** denote a collection of documents for sentiment classification. Each document **d** $\in$ D is first tokenized into a word sequence with maximum length **n**, then the word embeddings $\mathbf{w^i}$ of these sequence are jointly employed to represent the document $\mathbf{d = \{w^1, w^2, \ldots, wi, \ldots, w^n\}(i \in 1, 2, \ldots, n)}$.

To inject the sensory features into the neural framework, we apply a word embedding layer. We use perceptual vector to encode the input document **d** in Fig. 1. For sensory learning, each word embedding is first factorized into 6 embedding factors, each of which will be adapted to capture sensory semantics over each sensory dimension (i.e., auditory, gustatory, haptic, interoceptive, olfactory, and visual domains) through training. In other words, the sensory representation of the word is directly expressed in lexicon-level. For the other two groups, we choose the more traditional embedding models. One is based on the GloVe [15] pre-training model for word embedding, and the other uses random embedding.

LSTM encoder adopts the mode of gated mechanism, including Input Gate, Forget Gate, and Output Gate. Its core is Cell State, which refers to the state of the cell used

[1] https://www.mturk.com

for information dissemination. Memory Cell accepts two inputs: the output value of the last time h_{t-1} and the input value at this time x_t. Send these two parameters into the Forget Gate first, and get the information with less weight f_t. f_t is the information that we want to forget.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

Then we can enter the Input Gate and get the information to be updated i_t, which is the information with larger weight compared with the previous cell. Input Gate also gets the cell state $\tilde{C}_t$ at this time.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$
$$\tilde{C}_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3)$$

With Forget Gate and Input Gate, we can update cell state C_t.

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (4)$$

Finally, the output values of the Forget Gate and Input Gate are combined to calculate the output signal h_t, where at the same time, h_t is also the input signal of the next time.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$
$$h_t = o_t \cdot tanh(C_t) \quad (6)$$

Ensemble learning. In the final classification layer, we should fuse the features from both the two regular representations learning branches and perceptual learning branch. To that end, the three branches first handle the sentiment classification tasks separately. The learned document representations from each branch are first fed into the dense layers with sigmoid activation to yield the sentiment prediction. The document vector v is a high level representation of the document and can be used as features for document classification:

$$p = sigmoid(W_c v + b_c) \quad (7)$$

We use the Binary Cross-Entropy of the correct labels as training loss:

$$L = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i)) \quad (8)$$

So far, we have obtained the sentiment prediction of three kinds of word embedding sources. Finally, the three branches are fused to the final prediction result through voting.

4 Experiment results

Performance evaluations are conducted on two datasets and our model is compared with a series of commonly used baseline methods.

4.1 Datasets

The two benchmarking datasets include movie reviews from IMDB, and a toxic comment dataset [16]. For IMDB dataset, the reviews are collected from IMDB website, where each review is associated with a binary sentiment label indicating positive or negative polarity. The toxic comment dataset which is widely used for toxic detection and Kaggle competition founded by Jigsaw and Google. This dataset is seven classes and the types of toxicity are: toxic, severe toxic, obscene, threat, insult, and identity hate. We merge six types of toxic as positive, no toxic as negative.

Table 1 shows the statistics of the five datasets for experiments including the number of the train set, the number of test set, the ratio of classes, the number of vocabulary and the number of classes. They exhibit different characteristics, allowing the model evaluations in various scenarios.

Dataset	N_{train}	N_{test}	$R_{classes}$(pos:neg)	N_{voc}	C
IMDB	25,000	25,000	25,000 : 25,000	88582	2
Toxic Dataset	119,621	39,874	16,149 : 143,346	166465	2

Table 1: Statistics of the two benchmark datasets

4.2 Baseline Methods

To examine the effects of perceptual knowledge in experimental comparisons, three baseline systems are used in the evaluation in table 2.

	IMDB			Toxic Dataset		
Model	Acc	F1	AUC	Acc	F1	AUC
LSTM-Random	0.865	0.856	0.857	0.962	0.803	0.868
LSTM-GloVe	0.880	0.882	0.883	0.964	0.805	0.871
LSTM-Perceptual	0.870	0.869	0.858	0.964	0.803	**0.877**
Ensemble Model	**0.892**	**0.891**	**0.889**	**0.966**	**0.814**	0.875

Table 2: Evaluation of different methods; best result in accuracy is marked in bold

Their main settings are described as follows:

• **LSTM-Random** takes 128 dimension random embeddings to represent the document vectors and use LSTM as classifier.

• **LSTM-GloVe** takes the word embeddings of GloVe pre-train model to represent the document vectors and uses LSTM as classifier.

• **LSTM-Perceptual** takes the word embeddings of perceptual six-dimension-vector model to represent the document vectors and uses LSTM as classifier.

• **Ensemble Model** gets the final prediction is made by integrating the above three baseline models.

4.3 Performance Evaluation

The performance measures we use include Accuracy (ACC), F1-score (F1) and Area Under Curve (AUC) for our model. Table 2 shows the result of our experiment. In these two groups of sentiment classification tasks, Ensemble Model exhibits the best accuracy, but the accuracy of LSTM-Perceptual is less than LSTM-GloVe. This proves that for some vocabularies in some documents, perceptual information representation is helpful for sentiment classification. But not all vocabularies are applicable. The ensemble Model also achieves the best performance at F1 and AUC in IMDB. But the results are with only a 0.2% accuracy improvement over other models in Toxic Dataset. One possible explanation is that the toxic dataset is extremely unbalanced. As shown in Table 1. The ratio of positive cases to negative cases is only 0.11:1. Therefore the accuracy is usually very high and will not be greatly improved. In this unbalanced data set, the value of F1 is particularly important. F1-score can show the actual ability of the model on the unbalanced data set. Another reason may be that there are limited connections between sensory words and toxic language behaviors. The ensemble Model exhibits the best accuracy and results with a 1% F1-score improvement over other models. The above data shows that sensory information representation can improve the performance of sentiment classification tasks.

4.4 Case Study

To explore how sensory modalities knowledge improves sentiment classification tasks, we explore the specific case study. First, we chose the word "sad," which ultimately represents negative. In the GloVe pre-training word embedding, another word that describes negative "gloomy" only has a similarity of 0.36 with "sad." In contrast, in our perceptual space vector model, "gloomy" and "sad" have a similarity of up to 0.99. This shows that in some cases, the perceptual vector model has a more beneficial effect on sentiment classification than the traditional methods and makes the classification results tend to be correct.

5 Conclusion

This paper proposes a novel cognition-grounded model to improve the neural sentiment analysis model through sensory features of lexical items from external resources. An ensemble model takes the combined effects of bi-directional feeling between the external knowledge and the writer to infer human affective knowledge. The ensemble model considers both textual and sensory features. Evaluations of benchmark datasets validate the effectiveness of our method in sentiment analysis and related tasks, as our method outperforms other baseline approaches that use local context information to build their neural models. Thus, sensory data containing significant affective meaning can be leveraged into affective analysis tasks. The ensemble mechanism can also be combined with other models to provide room for further improvement. An important finding of our work is that sensory rating data gives a better gain in capturing synonyms and is closely related concepts than original word embedding information. This explains the improved performance of the sentiment-related task. While in the toxic language detection task, our work has a relatively limited improvement.

Our work also indicates that the quality and scale of sensory lexicon greatly influence the ensemble model's effectiveness. We anticipate even more significant improvements with a larger scale of data in more fine-grained sensory information. Another future work is building a more deep-level fusing model by leveraging affective lexicon into both feature and model levels.

References

[1] Cambria, E., Olsher, D., & Rajagopal, D. (2014, June). SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. In *Twenty-eighth AAAI conference on artificial intelligence*.

[2] Chen, H., Sun, M., Tu, C., Lin, Y., & Liu, Z. (2016, November). Neural sentiment classification with user and product attention. In *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 1650-1659).

[3] Dong, R., O'Mahony, M. P., Schaal, M., McCarthy, K., & Smyth, B. (2013, October). Sentimental product recommendation. In *Proceedings of the 7th ACM Conference on Recommender Systems* (pp. 411-414).

[4] Fang, H., Xu, G., Long, Y., & Tang, W. (2022). An Effective ELECTRA-Based Pipeline for Sentiment Analysis of Tourist Attraction Reviews. *Applied Sciences*, *12*(21), 10881.

[5] Kim Y. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Doha, Qatar (2014), pp. 1746-1751

[6] Lin, Z., Liang, B., Long, Y., Dang, Y., Yang, M., Zhang, M., & Xu, R. (2022, October). Modeling Intra-and Inter-Modal Relations: Hierarchical Graph Contrastive Learning for Multimodal Sentiment Analysis. In *Proceedings of the 29th International Conference on Computational Linguistics* (pp. 7124-7135).

[7] Long, Y., Lu, Q., Xiang, R., Li, M., & Huang, C. R. (2017, September). A cognition based attention model for sentiment analysis. In *Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 462-471).

[8] Long, Y., Ma, M., Lu, Q., Xiang, R., & Huang, C. R. (2018, October). Dual Memory Network Model for Biased Product Review Classification. In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (pp. 140-148).

[9] Long, Y., Xiang, R., Lu, Q., Huang, C. R., & Li, M. (2019). Improving attention model based on cognition grounded data for sentiment analysis. *IEEE transactions on affective computing*, *12*(4), 900-912.

[10] Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior research methods, 41*(2), 558-564.

[11] Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The Lancaster sensorimotor norms: Multidimensional measures of perceptual and action strength for 40,000 English words. Behavior research methods, 52(3), 1271-1291. doi:10.3758/s13428-019-01316-z

[12] Malins, S., Figueredo, G., Jilani, T., Long, Y., Andrews, J., Rawsthorne, M., ... & Moghaddam, N. (2022). Developing an Automated Assessment of In-session Patient Activation for Psychological Therapy: Codevelopment Approach. *JMIR Medical Informatics*, *10*(11), e38168.

[13] Nazir, A., Rao, Y., Wu, L., & Sun, L. (2020). Issues and challenges of aspect-based sentiment analysis: a comprehensive survey. *IEEE Transactions on Affective Computing*.

[14] Winter, B. (2016). Taste and smell words form an affectively loaded and emotionally flexible part of the English lexicon. *Language, Cognition and Neuroscience, 31*(8), 975-988.

[15] Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, A meeting of SIGDAT, a Special Interest Group of the ACL, Doha, Qatar, pp. 1532-1543.

[16] Wulczyn, E., Thain, N., Dixon, L.: Ex machina: Personal attacks seen at scale. In:Proceedings of the 26th international conference on world wide web, (2017). 1391-1399.

[17] Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016, June). Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies* (pp. 1480-1489).

[18] Zhao, Q., Huang, C.-R., & Lee, Y.-M. S. (2018). From linguistic synaesthesia to embodiment: Asymmetrical representations of taste and smell in Mandarin Chinese. In Wu, Y., Hong, J.-F. & Su, Q. (eds.): *The 18th Chinese Lexical Semantics Workshop (CLSW-2017), LNAI 10709*, Springer (pp. 406-413).