Enhancing Cognitive Models of Emotions with Representation Learning

Yuting Guo
Computer Science
Emory University
Atlanta GA 30322, USA
`yuting.guo@emory.edu`

Jinho D. Choi
Computer Science
Emory University
Atlanta GA 30322, USA
`jinho.choi@emory.edu`

Abstract

We present a novel deep learning-based framework to generate embedding representations of fine-grained emotions that can be used to computationally describe psychological models of emotions. Our framework integrates a contextualized embedding encoder with a multi-head probing model that enables to interpret dynamically learned representations optimized for an emotion classification task. Our model is evaluated on the Empathetic Dialogue dataset and shows the state-of-the-art result for classifying 32 emotions. Our layer analysis can derive an emotion graph to depict hierarchical relations among the emotions. Our emotion representations can be used to generate an emotion wheel directly comparable to the one from Plutchik's model, and also augment the values of missing emotions in the PAD emotional state model.

1 Introduction

Emotion classification has been extensively studied by many disciplines for decades (Spencer, 1895; Lazarus and Lazarus, 1994; Ekman, 1999). Two main streams have been developed for this research: one is the discrete theory that tries to explain emotions with basic and complex categories (Plutchik, 1980; Ekman, 1992; Colombetti, 2009), and the other is the dimensional theory that aims to conceptualize emotions into a continuous vector space (Russell and Mehrabian, 1977; Watson and Tellegen, 1985; Bradley et al., 1992). Illustration of human emotion however is often subjective and obscure in nature, leading to a long debate among researchers about the "correct" way of representing emotions (Gendron and Feldman Barrett, 2009).

Representation learning has made remarkable progress recently by building neural language models on large corpora, which have substantially improved the performance on many downstream tasks (Peters et al., 2018; Devlin et al., 2019; Yang et al., 2019; Liu et al., 2019; Joshi et al., 2020). Encouraged by this rapid progress along with an increasing

interest of interpretability in deep learning models, several studies have attempted to capture various knowledge encoded in language (Adi et al., 2017; Peters et al., 2018; Hewitt and Manning, 2019), and shown that it is possible to learn computational representations through distributional semantics for abstract concepts. Inspired by these prior studies, we build a deep learning-based framework to generate emotion embeddings from text and assess its ability of enhancing cognitive models of emotions. Our contributions are summarized as follows:[1]

- To develop a deep probing model that allows us to interpret the process of representation learning on emotion classification (Section 3).

- To achieve the state-of-the-art result on the Empathetic Dialogue dataset for the classification of 32 emotions (Section 4).

- To generate emotion representations that can derive an emotion graph, an emotion wheel, as well as fill the gap for unexplored emotions from existing emotion theories (Section 5).

2 Related Work

Probing models are designed to construct a probe to detect knowledge in embedding representations. Peters et al. (2018) used linear probes to examine phrasal information in representations learned by deep neural models on multiple NLP tasks. Tenney et al. (2019) proposed an edge probing model using a span pooling to analyze syntactic and semantic relations among words through word embeddings. Hewitt and Manning (2019) constructed a structural probe to detect the correlations among word pairs to predict their latent distances in dependency trees. As far as we can tell, our work is the first to generate embeddings of fine-grained emotions from text and apply them to well-established emotion theories.

---

[1]All our resources including source codes and models are available at `https://github.com/emorynlp/CMCL-2021`.
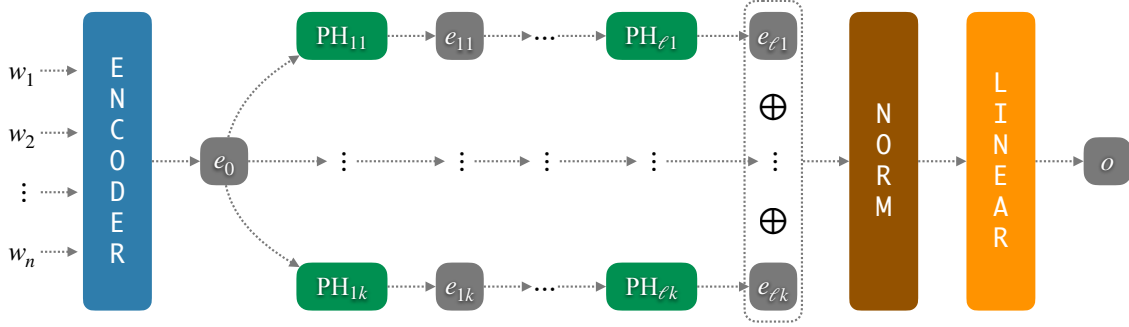
Figure 1: The overview of our deep learning-based multi-head probing model.

NLP researchers have produced several corpora for emotion detection including FriendsED (Zahiri and Choi, 2018), EmoInt (Mohammad et al., 2017), EmoBank (Buechel and Hahn, 2017), and Daily-Dialogs (Li et al., 2017), all of which are based on coarse-grained emotions with at most 7 categories. For a more comprehensive analysis, we adapt the Empathetic Dialogue dataset based on fine-grained emotions with 32 categories (Rashkin et al., 2019).

3 Multi-head Probing Model

We present a multi-head probing model allowing us to interpret how emotion embeddings are learned in deep learning models. Figure 1 shows an overview of our probing model. Let $W = \{w_1, \ldots, w_n\}$ be an input document where w_i is the i'th token in the document. W is first fed into a contextualized embedding encoder that generates the embedding $e_0 \in \mathbb{R}^{d_0}$ representing the entire document. The document embedding e_0 is then fed into multiple probing heads, $\text{PH}_{11}, \ldots, \text{PH}_{1k}$, that generate the vectors $e_{1j} \in \mathbb{R}^{d_1}$ comprising features useful for emotion classification ($j \in [1, k]$). The probing heads in this layer are expected to capture abstract concepts (e.g., positive/negative, intense/mild).

Each vector e_{1j} is fed into a sequence of probing heads where the probing head PH_{ij} is defined $\text{PH}_{ij}(e_{hj}) \rightarrow e_{ij}$ ($i \in [2, \ell], j \in [1, k], h = i - 1$). The feature vectors $e_{\ell*}$ from the final probing layer are expected to learn more fine-grained concepts (e.g., ashamed/embarrassed, hopeful/anticipating). $e_{\ell*}$ are concatenated and normalized to $g_\ell \in \mathbb{R}^{d_\ell \cdot k}$ and fed into a linear layer that generates the output vector $o \in \mathbb{R}^m$ where m is the total number of emotions in the training data. It is worth mentioning that every probing sequence finds its own feature combinations. Thus, each of $e_{\ell*}$ potentially represents different concepts in emotions, which allow us to analyze concept compositions of these emotions empirically derived by this model.

4 Experiments

4.1 Contextualized Embedding Encoder

For all experiments, BERT (Devlin et al., 2019) is used as the contextualized embedding encoder for our multi-head probing model in Section 3. BERT prepends the special token CLS to the input document W such that $W' = \{\text{CLS}\} \oplus W$ is fed into the ENCODER in Figure 1 instead, which generates the document embedding e_0 by applying several layers of multi-head attentions to CLS along with the other tokens in W (Vaswani et al., 2017).[2]

4.2 Dataset

Although several datasets are available for various types of emotion detection tasks (Section 2), most of them are annotated with coarse-grained labels that are not suitable to make a comprehensive analysis of emotions learned by deep learning models.

	TRN	DEV	TST	ALL
C	19,533	2,770	2,547	24,850
L	18.2 (±10.4)	19.6 (±11.4)	23.0 (±12.5)	18.9 (±10.8)

Table 1: Statistics of the Empathetic Dialogue dataset. TRN/DEV/TST: training/development/test set. C: # of documents, L: average # of tokens and its standard deviation in each document.

To demonstrate the impact of our probing model, the Empathetic Dialogue dataset is selected, that is labeled with 32 emotions on ≈25K conversations related to daily life, each of which comes with an emotion label, a situation described in text that can reflect the emotion (e.g., Proud → "*I finally got that promotion at work!*"), and a short two-party dialogue generated through MTurk that simulates a conversation about the situation (Rashkin et al., 2019). For our experiments, only the situation parts are used as input documents.

---

[2]Details about the experimental settings are provided in Section A.1.
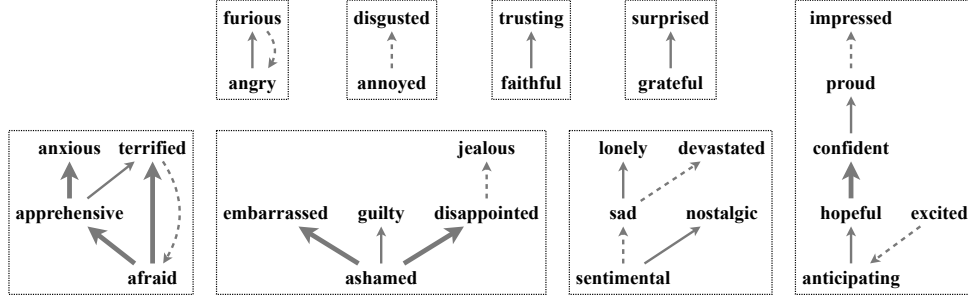
142

Figure 2: The overview of our deep learning-based multi-head probing model.

4.3 Results

Several multi-head probing models are developed by varying the number of probing layers and the dimension of feature vectors to find the most effective model for interpretation. For all models, a linear layer is used for every probing head such that $\mathrm{PH}_{i*}(e_{h*}) \to w \cdot e_{h*} = e_{i*}$, where $e_{h*} \in \mathbb{R}^{1 \times d_h}$, $w \in \mathbb{R}^{d_h \times d_i}$, $e_{i*} \in \mathbb{R}^{1 \times d_i}$. The dimension of the document embedding d_0 is set to 768 for all models as configured by the pretrained BERT model.

k	128:64:32	64:32	32
2	56.9 ($\pm$0.4)	57.1 ($\pm$0.5)	56.9 ($\pm$0.5)
4	57.5 ($\pm$0.4)	58.1 ($\pm$0.5)	**57.8** ($\pm$0.5)
8	**57.8** ($\pm$0.8)	**58.2** ($\pm$0.5)	57.6 ($\pm$0.1)
16	57.2 ($\pm$0.3)	57.6 ($\pm$0.4)	57.7 ($\pm$0.6)
32	57.2 ($\pm$0.9)	57.3 ($\pm$0.4)	57.5 ($\pm$0.7)
64	56.8 ($\pm$0.6)	57.2 ($\pm$0.3)	57.4 ($\pm$0.4)

Table 2: Average accuracies and standard deviations on the test set. k: total # of feature vectors in each layer, i'th # in each column delimited by colons is the dimension of the feature vectors in the i'th probing layer.

Table 2 shows the results achieved by all models; every model is trained 3 times and the average accuracy and its standard deviation is reported. The baseline BERT model using no probing, that is to feed e_0 directly into the linear layer, is also built for comparison, showing a significantly higher accuracy of 57.6% ($\pm$0.02) than the previously reported state-of-the-art of 48% by Rashkin et al. (2019). The best result is achieved by the 2-layer probing model with 8 feature vectors, showing the accuracy of 58.2% ($d_1 = 64, d_2 = 32, k = 8$).

5 Analysis

5.1 Layer-wise Analysis

To analyze which emotional concepts are embedded in each probing layer (Section 3), we train a logistic regression model on the concatenated vector of $(e_{i1} \oplus \cdots \oplus e_{ik})$ for each layer ℓ_i with the same configuration used for the 3-layer model,

128:64:32 (Table 2), and tested on the development set. For each pair of adjacent layers (ℓ_i, ℓ_j) where $j = i+1$ and $1 \leq i \leq 2$, we measure the likelihood $H_{ij}(s, t)$ of those layers classifying each emotion s as every other emotion t as follows:

$$H_{ij}(s, t) = L(s, t) - L(t, s)$$
$$L(e_g, e_p) = \ell_j(e_g, e_p) - \ell_i(e_g, e_p)$$

where $\ell_*(e_g, e_p)$ is the proportion of the documents whose gold labels are e_g but predicted as e_p by the model trained on the layer ℓ_*. If $L(s, t) > 0$, it means that the higher layer ℓ_j tends to predict s as t more than the lower layer ℓ_i. $L(t, s) > 0$ implies the opposite, and is used as a penalty term to get a more reliable measurement of how much the higher layer is confused s for t than the lower layer.

The results are illustrated in Figure 2, where arrows pointing from one emotion s to another emotion t indicate $H_{ij}(s, j) \geq 2$. The dashed arrows and thin solid arrows correspond to the confusion likelihoods of $H_{12}(s, j)$ and $H_{23}(s, j)$ respectively, and the thick solid arrows reflect the likelihoods in those two metrics. Most emotion pairs point from coarse-grained emotions to fine-grained emotions (e.g., *angry* $\to$ *furious*, *sentimental* $\to$ *nostalgic*) except for a few pairs (*excited* $\to$ *anticipating*), implying that higher probing layers tend to learn more finer-grained emotions that lower layers.

5.2 Generation of Emotion Wheel

Plutchik (1980) introduced the emotion wheel by selecting a reference emotion and arranging others on a circle where the angles are determined by manually assessed similarities between emotion pairs. Inspired by this work, we derive an emotion wheel by creating emotion embeddings and representing each complex emotion as a weighted sum of two basic emotions. Given an emotion e and a set of documents $\mathcal{D}_e$ whose gold labels are e in the DEV set, the embedding of e can be derived as follows,

where g_ℓ^d is the normalized vector in Section 3 for d.

$$r_e = \frac{1}{|\mathcal{D}_e|} \sum_{\forall d \in \mathcal{D}_e} g_\ell^d \qquad (1)$$

For each complex emotion c, its combinatory basic emotion pair (b_i, b_j) and the weight $w \in [0.1, 0.9]$ are founded as follows (r_* is the embedding of b_*):

$$r_{i,j,w} = w \cdot r_i + (1 - w) \cdot r_j$$
$$(b_i, b_j, w) = \underset{\forall i, \forall j, \forall w}{\arg \max}[\text{cosine\_sim}(r_{i,j,w}, c)] \quad (2)$$

Figure 3 depicts the emotion wheel auto-generated by our framework; 8 basic emotions are displayed on the outer circle and complex emotions are displayed on the edges between those basic emotions where the dot scales are proportional to the cosine_sims in Eq (2).[3] Although the only manual part in this wheel is the selection of those basic emotions from Plutchik (1980), it is compatible to the original emotion wheel in Section A.2 and finds even more relations such as *Excited = Anticipating + Joyful*, *Lonely = Sad + Afraid*, and *Grateful = Trusting* + Joyful.
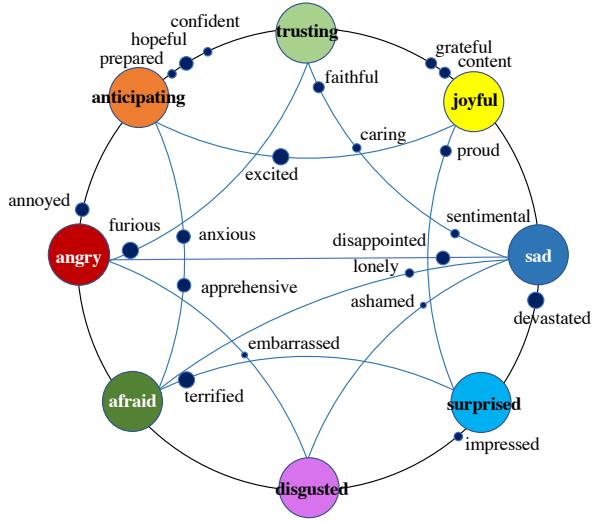

Figure 3: Emotion wheel auto-derived by our approach.

5.3 Augmentation of PAD Model

Russell and Mehrabian (1977) presented the PAD model suggesting that emotions can be denoted by 3 dimensions of pleasure, arousal, and dominance. To verify whether our representations can capture emotional concepts similar to the PAD model, we train a regression model per dimension that takes the emotion embeddings from Eq (1) and learns the corresponding PAD values in Section A.3 manually assessed by Russell and Mehrabian (1977).

Note that the original PAD model provides the PAD values for only 22 emotions. Given the 3 regression models trained on those 22 emotions, we are able to predict the PAD values for the other 10 emotions missing from the original model.[4] Figure 4 shows the 2D plot of the PA values predicted by our regression models for Pleasure and Arousal, where the 10 emotions, whose PAD values are newly discovered by our models, are indicated with the red labels.[5] It is exciting to see that the newly discovered emotions blend well in this plot (e.g., *anticipating* in between *anxious* and *excited*). Similar emotions are closer in this space (e.g., *sentimental / nostalgic*, *trusting / faithful / confident*), implying the robustness of the predicted values. Notice that the P value of *nostalgic* is predicted as positive, which is understandable because *nostalgic* is related to a memory with happy personal associations; thus, it is found to be positive by distributional semantics.
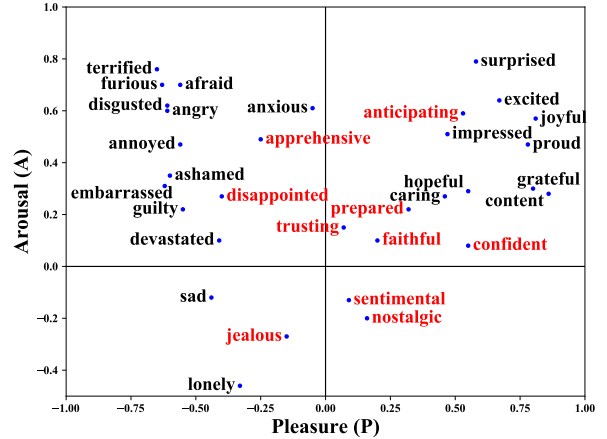

Figure 4: The 2D plot from the PAD values of 32 emotions predicted by our regression models.

6 Conclusion

This paper presents a multi-head probing model to derive emotion embeddings from neural model interpretation. Our model is applied to an emotion detection task and shows a state-of-the-art result. These emotion embeddings can derive an emotion graph, depicting how abstract concepts are learned in neural models, and an emotion wheel and PAD values, verifying their potential of augmenting cognitive models for more diverse groups of emotions that have not been explored by cognitive theories.

---

[3]3 complex emotions whose cosine similarity scores are less than 0.1 are omitted in Figure 3: *guilty, jealous, nostalgic*.

[4]Section A.3 provides configurations for all three models.

[5]The 3D plot including the dominance values is in Section A.3.

References

Yossi Adi, Einat Kermany, Yonatan Belinkov, Ofer Lavi, and Yoav Goldberg. 2017. Fine-grained Analysis of Sentence Embeddings Using Auxiliary Prediction Tasks. *5th International Conference on Learning Representations*.

Margaret M. Bradley, Mark K. Greenwald, Margaret C. Petry, and Peter J. Lang. 1992. Remembering pictures: Pleasure and arousal in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(2):379–390.

Sven Buechel and Udo Hahn. 2017. EmoBank: Studying the Impact of Annotation Perspective and Representation Format on Dimensional Emotion Analysis. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 578–585, Valencia, Spain. Association for Computational Linguistics.

Giovanna Colombetti. 2009. From affect programs to dynamical discrete emotions. *Philosophical Psychology*, 22(4):407–425.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186.

Paul Ekman. 1992. An Argument for Basic Emotions. *Cognition & Emotion*, 6(3/4):169–200.

Paul Ekman. 1999. Basic Emotions. *Handbook of Cognition and Emotion*, 98(45-60):16.

Maria Gendron and Lisa Feldman Barrett. 2009. Reconstructing the Past: A Century of Ideas About Emotion in Psychology. *Emotion Review*, 1(4):316–339.

John Hewitt and Christopher D Manning. 2019. A Structural Probe for Finding Syntax in Word Representations. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4129–4138.

Mandar Joshi, Danqi Chen, Yinhan Liu, Daniel S. Weld, Luke Zettlemoyer, and Omer Levy. 2020. SpanBERT: Improving Pre-training by Representing and Predicting Spans. *Transactions of the Association for Computational Linguistics 2020*.

Richard S Lazarus and Bernice N Lazarus. 1994. *Passion and Reason: Making Sense of Our Emotions*. Oxford University Press, USA.

Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017. DailyDialog: A Manually Labelled Multi-turn Dialogue Dataset. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 986–995, Taipei, Taiwan. Asian Federation of Natural Language Processing.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv*, 1907(11692).

Saif M. Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. 2017. Stance and Sentiment in Tweets. *ACM Transactions on Internet Technology*, 17(3):1–23.

Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep Contextualized Word Representations. In *Proceedings of the 2018 North American Chapter of the Association for Computational Linguistics*, pages 2227–2237.

Robert Plutchik. 1980. A General Psychoevolutionary Theory of Emotion. In *Theories of Emotion*, pages 3–33. Elsevier.

Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2019. Towards Empathetic Open-domain Conversation Models: A New Benchmark and Dataset. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5370–5381, Florence, Italy. Association for Computational Linguistics.

James A Russell and Albert Mehrabian. 1977. Evidence for A Three-Factor Theory of Emotions. *Journal of Research in Personality*, 11(3):273–294.

Herbert Spencer. 1895. *The Principles of Psychology*, volume 1. Appleton.

Ian Tenney, Patrick Xia, Berlin Chen, Alex Wang, Adam Poliak, R. Thomas McCoy, Najoung Kim, Benjamin Van Durme, Samuel R. Bowman, Dipanjan Das, and Ellie Pavlick. 2019. What Do You Learn From Context? Probing for Sentence Structure in Contextualized Word Representations. In *9th International Conference on Learning Representations*, pages 55–65.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *31st Conference on Neural Information Processing Systems*.

David Watson and Auke Tellegen. 1985. Toward a Consensual Structure of Mood. *Psychological Bulletin*, 98(2):219–235.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. 2019. XLNet: Generalized Autoregressive Pretraining for Language Understanding. In *Advances in Neural Information Processing Systems 32*, pages 5753–5763.

Sayyed Zahiri and Jinho D. Choi. 2018. Emotion Detection on TV Show Transcripts with Sequence-based Convolutional Neural Networks. In *Proceedings of the AAAI Workshop on Affective Content Analysis*, AFFCON'18, pages 44–51, New Orleans, LA.

A Appendix

A.1 Experimental Settings

The BERT model used in our experiment is *BERT-base*, and Table 3 shows the hyperparameters used to develop the models in Table 2.

Hyperparameter	Value
n: max document length	128
m: number of classes	32
k: number of feature vectors in each layer	8
d_0: dimension of the feature vector e_0	768
batch size	32
learning rate	5e-5

(a) Shared hyperparameters.

	128:64:32	64:32	32
l: # of probing layers	3	2	1
d_1: dimension of e_1	128	64	32
d_2: dimension of e_2	64	32	-
d_3: dimension of e_3	32	-	-

(b) Model-specific hyperparameters.

Table 3: Hyperparameter configurations for all models.

A.2 Plutchik's Emotion Wheel

The emotion wheel described in Section 5.2 is inspired by Plutchik (1980) which proposed the eight basic emotions that can constitute other complex emotions through various combinations shown by the emotion wheel in Figure 5, where emotions displayed on the edges are the compositions of those two basic emotions. As can be seen, our derived emotion wheel has some identical emotion relations as the Plutchik's emotion wheel such as *Hope* = *Anticipation* + *Trust*, *Anxiety* = *Anticipation* + *Fear*, and *Sentimentality* = *Trust* + *Sadness*. It suggests the robustness of the emotion wheel derived by the proposed method in Section 5.2.

A.3 Russell and Mehrabian's PAD Model

All regression models in Section 5.3 are based on 2-layer multilayer perceptron using the mean square error (MSE) loss, including a hidden layer with the `ReLU` activation and an output layer with the `Tanh` activation. The hidden layer dimension is 128, and the dropout rate is 0.3, and early stopping is applied to avoid overfitting. The MSE losses of the three regression models to predict the Pleasure (P), Arousal (A), and Dominance (D) values are 0.028, 0.019, and 0.016, respectively. Table 4 describes the original PAD values of the 22 emotions from Russell and Mehrabian (1977), and Figure 6 shows the 2D plot from the PAD values of those 22 emotions. Table 5 describes the PAD values predicted
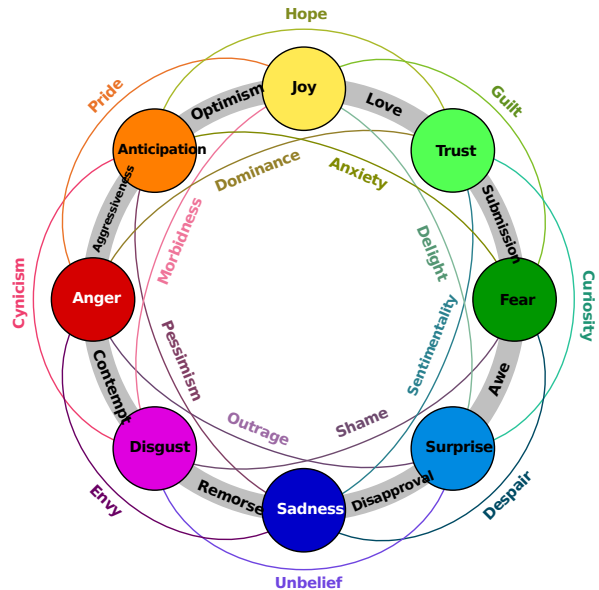

Figure 5: Emotion wheel proposed by Plutchik (1980).

by our regressions models, which are plotted in Figure 4. Finally, Figure 7 plots those predicted PAD values in the 3D space to depict the dominance values with respect to the other two PA dimensions. By comparing the PAD values of 22 emotions in Table 4 and Table 5, most of the predicted values are close to their gold values. Also, we can observe that the predicted values of some newly discovered emotions are consistent with our perception of emotions. For example, *Anticipating* is very close to *Hope* in terms of pleasure but with higher intensity.

Emotion	Pleasure	Arousal	Dominance
afraid	-0.64	0.6	-0.43
angry	-0.51	0.59	0.25
annoyed	-0.28	0.17	0.04
anxious	0.01	0.59	-0.15
ashamed	-0.57	0.01	-0.34
caring	0.64	0.35	0.24
content	0.86	0.2	0.62
devastated	0.14	0.45	-0.24
disgusted	-0.6	0.35	0.11
embarrassed	-0.46	0.54	-0.24
excited	0.62	0.75	0.38
furious	-0.44	0.72	0.32
grateful	0.64	0.16	-0.21
guilty	-0.57	0.28	-0.34
hopeful	0.51	0.23	0.14
impressed	0.41	0.3	-0.32
joyful	0.76	0.48	0.35
lonely	-0.66	-0.43	-0.32
proud	0.77	0.38	0.65
sad	-0.64	-0.27	-0.33
surprised	0.4	0.67	-0.13
terrified	-0.62	0.82	-0.43

Table 4: The original PAD values of 22 emotions provided by Russell and Mehrabian (1977).
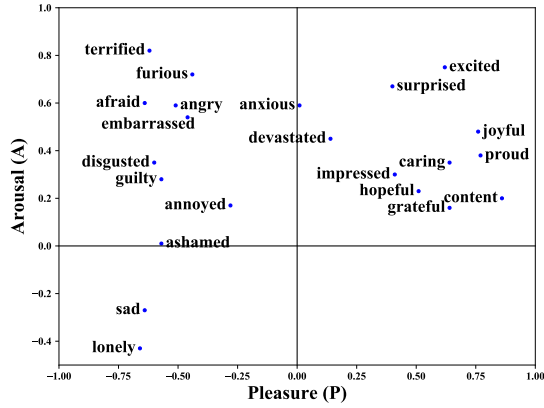
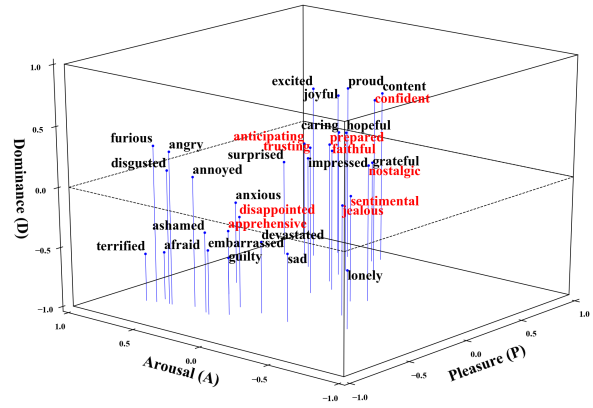Figure 6: The 2D plot from the PAD values in Table 4.

Figure 7: The 3D plot from the PAD values in Table 5.

Emotion	Pleasure	Arousal	Dominance
afraid	-0.56	0.7	-0.6
angry	-0.61	0.6	0.28
annoyed	-0.56	0.47	0.09
anticipating	**0.53**	**0.59**	**0.03**
anxious	-0.05	0.61	-0.31
apprehensive	**-0.25**	**0.49**	**-0.46**
ashamed	-0.6	0.35	-0.33
caring	0.46	0.27	0.22
confident	**0.55**	**0.08**	**0.51**
content	0.86	0.28	0.44
devastated	-0.41	0.1	-0.4
disappointed	**-0.4**	**0.27**	**-0.24**
disgusted	-0.61	0.62	0.12
embarrassed	-0.62	0.31	-0.46
excited	0.67	0.64	0.45
faithful	**0.2**	**0.1**	**0.18**
furious	-0.63	0.7	0.31
grateful	0.8	0.3	-0.14
guilty	-0.55	0.22	-0.52
hopeful	0.55	0.29	0.19
impressed	0.47	0.51	-0.06
jealous	**-0.15**	**-0.27**	**-0.08**
joyful	0.81	0.57	0.37
lonely	-0.33	-0.46	-0.51
nostalgic	**0.16**	**-0.2**	**0.14**
prepared	**0.32**	**0.22**	**0.17**
proud	0.78	0.47	0.46
sad	-0.44	-0.12	-0.43
sentimental	**0.09**	**-0.13**	**-0.11**
surprised	0.58	0.79	-0.19
terrified	-0.65	0.76	-0.6
trusting	**0.07**	**0.15**	**0.23**

Table 5: The PAD values of 32 emotions predicted by our regression models. The 10 emotions that are missing from the original work in Table 4 are indicated with bold font.

The weight indicates how much each basic emotion in the pair contributes to the complex emotion and can be interpreted in a proportional manner. For example, *Annoyed* can be composed of 90% *Angry* and 10% *Anticipating*.

c	b_i	b_j	w	cos
annoyed	angry	anticipating	0.9	0.80
anxious	anticipating	afraid	0.5	0.79
apprehensive	anticipating	afraid	0.3	0.76
ashamed	sad	disgusted	0.6	0.17
caring	trusting	sad	0.5	0.28
confident	anticipating	trusting	0.5	0.31
content	joyful	trusting	0.9	0.63
devastated	surprised	sad	0.1	0.93
disappointed	sad	angry	0.7	0.64
embarrassed	disgusted	angry	0.5	0.13
excited	anticipating	joyful	0.5	0.95
faithful	trusting	sad	0.9	0.59
furious	angry	trusting	0.9	0.98
grateful	joyful	trusting	0.8	0.56
guilty	trusting	sad	0.1	0.07
hopeful	anticipating	trusting	0.8	0.67
impressed	surprised	disgusted	0.9	0.40
jealous	disgusted	angry	0.3	0.02
lonely	afraid	sad	0.2	0.33
nostalgic	anticipating	joyful	0.1	0.04
prepared	anticipating	trusting	0.9	0.31
proud	joyful	surprised	0.9	0.45
sentimental	trusting	sad	0.1	0.33
terrified	afraid	surprised	0.9	0.98

Table 6: The combinatory basic emotion pairs for each complex emotion.

A.4 Combinatory Emotions Details

In Section 5.2, we propose a framework to find the combinatory basic emotion pairs for each complex emotion by calculating a weighted sum vector of two basic emotion embeddings. Table 6 lists the basis emotion pairs, weights, and cosine similarity for 24 complex emotions derived by our framework.