A Mapping on Current Classifying Categories of Emotions Used in Multimodal Models for Emotion Recognition

Ziwei Gong [1] Xinyi Hu [2] Muyin Yao [3] Xiaoning Zhu [4] Julia Hirschberg [1]

[1]Columbia University [2]Boston University [3]Tufts University [4]JYLLink Co., Ltd.

Motivation

In Emotion Detection within Natural Language Processing and related multimodal research, the growth of datasets and models has led to a challenge: disparities in emotion classification methods. The lack of commonly agreed upon conventions on the classification of emotions creates boundaries for model comparisons and dataset adaptation. In this paper, we compare the current classification methods in recent models and datasets and propose a valid method to combine different emotion categories. Our proposal arises from experiments across models, psychological theories, and human evaluations, and we examined the effect of proposed mapping on models.

Mapping Method

Common Emotions: Emotions shared by both categories remain unaltered. Although these emotions might have different definitions across theories, our sample annotation process suggests annotators seldom find them non-transferable. Considering the annotation process of large datasets, it is common that their annotators are asked to choose an emotion that best describes the current scene or utterance rather than strictly following the definition of that emotion.

Higher-Level Emotions: Emotions exclusive to higher-level categories are mapped based on past literature, often considering valence and arousal of various emotions. Valence measures the positiveness or negativity of an emotional stimulus; arousal measures the intensity of emotion. Emotions with comparable arousal and valence levels are more likely to be paired, contrasting with emotions that differ in these aspects.

Human Evaluations: When faced with tied choices, we conduct human evaluations on each theory to determine the best mapping choice in the situation of a tie.

The Classification for Surprise as Example

Surprise characterizes the feeling of shock due to perceiving things or experience out of expectation. To map surprise, we employed a bipolar model integrating valence and arousal dimensions. Russell introduced this model in 1977, with motivation as an initial component. Surprise may be considered a negative emotion, since previous studies associate surprise with a negative valence and high arousal levels. Based on Liu et al.'s research, high-arousal, low-valence emotions are akin to anger. However, the potential for positive valence-associated surprise introduces ambiguity in conversion, possibly favoring mapping to neutral.

We leverage biological distinctions between emotions as a reference. A recent study utilizing biomarkers to analyze EEG profiles across brain regions offers valuable findings. Among surprise-combined emotions, the spectral biomarker's mean differences (0.114) and the temporal biomarker's mean differences (0.058) are lowest for the neutral-surprise pairing.

Hence, both anger and neutral are considered possible mappings for surprise. To test this hypothesis, we implemented a program to convert surprise into anger and neutral. These converted emotions were mixed with randomly selected samples of other emotions. Annotators, at least two per data point, participated in the evaluation. Evaluation results favored the surprise-to-anger conversion, as it achieved higher accuracy. Hence, we map surprise to anger based on annotation outcomes.

Map analysis

Our analysis shows negative emotions are more finely categorized than positive or neutral ones, with 8 "negative", 3 "positive", and 3 "neutral" distribution among 14 categories. This may be due to dataset biases and psychological factors.

The imbalance in emotion mapping, especially for ambiguous emotions like surprise and trust, may reflect TV shows and media biases and categorization theories, yet still achieves acceptable evaluation scores.

Despite categorization challenges for emotions like surprise and trust, our method achieved acceptable evaluation scores, indicating a bias towards negative interpretation due to existing classification theories. Our method aligns emotions with the same names across classifications, despite slight differences, offering a standardized approach. Its effectiveness may vary with dataset diversity.

As an initial attempt to standardize emotion classification from a psychological perspective, our work seeks to encourage further research in resolving classification disparities.

Mapping Results

Propose the first complete mapping that connects different emotion categories for multimodal emotion recognition.

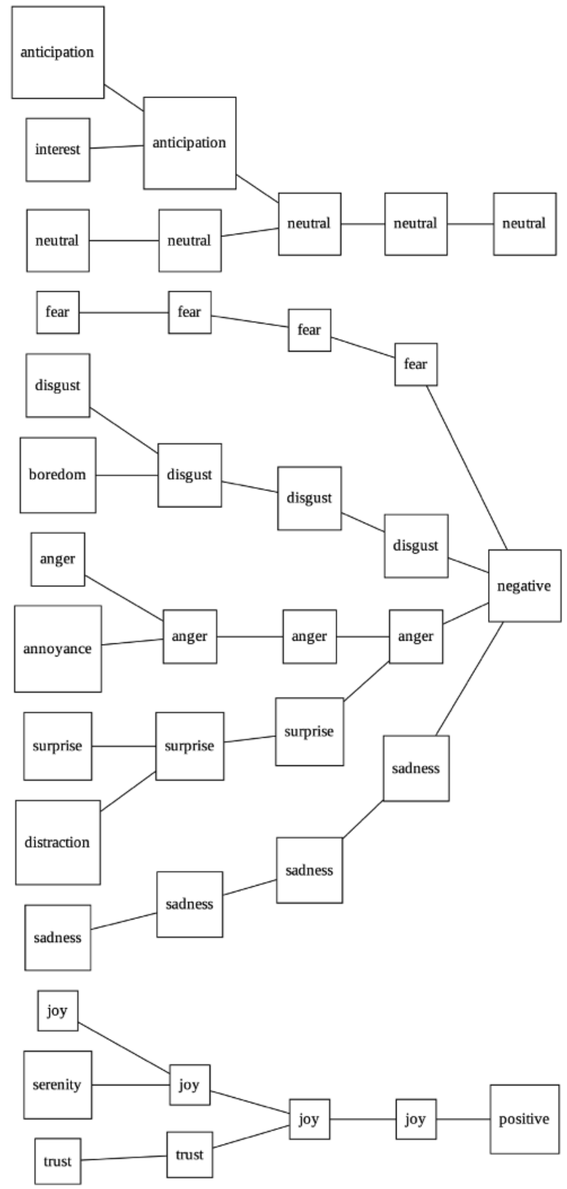

Figure 1. Mapping method in graph. This graph demonstrates how 14 fine-grained emotions, listed on the leftmost column, are mapped onto 9 primary emotions, Ekman's basic emotions, 6 emotions, and the 3 sentiments.

Mapping effects on ML Models

Emotion Category	3	6	7	9	14
MEmoR Accuracy	**0.924**	0.867	0.884	0.869	0.864
CNN Accuracy	**81.78**	65.39	65.28	-	-

Table 1. Experimental results from the MEmoR model and the CNN model. This table shows the overall accuracy of the models trained and tested on datasets reconstructed based on each 3 classification method. The MEmoR model uses visual, audio, textual features. In the CNN model, only visual information is used.

Figure 2. Contrast in attention heat maps across 9 random images: a CNN model trained on a 7-category dataset (left) vs. the same dataset categorized into 3 groups (right). Regions of high attention are shown in red.

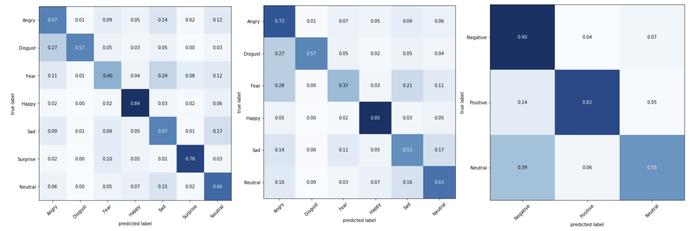

Figure 3. Confusion matrices generated by three CNN models trained on a dataset, all learning from the same set of pictures but with labels categorized into 7 (left), 6 (middle) and 3 categories (left). Columns represent the predicted label and rows represent the true label.

- From experiments we found models generally perform better when there are fewer emotion categories, meaning that more fine-grained emotions are more difficult for models to differentiate. (Table 1)
- We observe from Heat Map on CNN (vision) model that the attention of the model trained with more fine-grained emotions is more spread out through the face, compared to only focusing around the eye and mouth area. (Figure 2)
- From confusion matrics, improvement was mainly on the adjusted category. (Figure 3)

Using our mapping allows researchers to obtain larger and more flexible datasets for training, and to analyze models across different datasets.

Conclusions

In this paper, we propose the first complete mapping that connects different emotion categories for multimodal emotion recognition studies, and provide a study of the effect of using different emotion classification methods when training models. We attempt to bridge the different psychological emotion theories and lend them consistency in the computer science world. Moreover, using our mapping allows researchers to obtain a larger and more flexible dataset for training and testing and to analyze the model's ability to differentiate emotions using different emotion categories, as well as identify the best model across all datasets.

zg2272@columbia.edu