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

Ziwei Gong
Columbia University
zg2272@columbia.edu

Xinyi Hu
Boston University
xhu07@bu.edu

Muyin Yao
Tufts University
yaomuyin@gmail.com

Xiaoning Zhu
JYLlink Co., Ltd.
zhuxiaoning@jylink.com

Julia Hirschberg
Columbia University
julia@cs.columbia.edu

Abstract

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.

1 Introduction

Emotion recognition, as an essential ability for good interpersonal relations (Mancini et al., 2018), has long been a major subject in psychology, and for the last two decades has received increasing attention from the field of computer science, especially artificial intelligence (De Silva et al., 1997; Gong et al., 2023). Yet in this process a divergence has emerged from newly published datasets and models — the misalignment between different categories of emotions. To resolve such disparity between emotion datasets, we propose a psychology-based solution for computer scientists to solve the problem of misalignment in emotion classification datasets, which is caused by the independent nature of emotion classification theories.

In the field of psychology, there are many different theories on how to classify emotions focusing on different aspects. Various theories classify emotions based on different factors: Ekman's theory focuses on universal facial expressions (Ekman, 1992), comparing the facial expressions of westerners and Aboriginal residents of New Guinea; Plutchik's evolutionary perspective categorizes emotions into 8 primary emotions with 3 levels of intensity (Plutchik, 2001) based on the communication function of emotions; Barrett's (Wilson-Mendenhall et al., 2011) biological approach studies brain responses to emotions through intepreting EEGs and physiological changes (Hess, 2017); and emphasizing cultural influence, the constructionist theory adds social and linguistic elements to emotion understanding (Wilson-Mendenhall et al., 2011). These theories are independent but sometimes interconnected, providing a foundation for potential integration. Different theories are mostly considered to be independent theories of emotion, yet these classification approaches are often interconnected and sometimes built upon each other, providing a basis to connect them. However, few studies explore ways to connect or combine these different categorizations.

In computer science, researchers face the challenge of choosing an emotion theory when building datasets for emotion detection. Recent work in emotion classification has shifted towards using multimodal data sources like audio, video, and text (Poria et al., 2019; Shen et al., 2020), and some even explore incorporating additional factors like personality and social connections to leverage more information for deep learning models (Kahou et al., 2015). Due to varying annotation methods and mismatch in the set of labels, a model typically selects a single dataset for experiments, although more data could improve its performance. A significant issue arises from the lack of alignment in labeling schemas across datasets, making it challenging for models to leverage multiple datasets in supervised learning (Bostan and Klinger, 2018). This disparity results in a lack of cohesion in the literature, hinders direct performance comparisons, and complicates dataset combination and training. Since annotating such datasets is costly and time-consuming, a mapping method that can unify existing datasets could benefit the community. Currently, little research in both psychology and com-

Proceedings of The 18th Linguistic Annotation Workshop (LAW-XVIII), pages 19–28

puter science explores the relationship between different emotion categories. While there are studies mapping categorical emotions onto dimensional models (Hoffmann et al., 2012) and recent work inproviding more grounded emotion categories in Dutch (De Bruyne et al., 2020), the mapping between multiple categorical emotions, which creates misalignment in emotion datasets for machine learning, remains unstudied.

This paper aims to establish a valid mapping of emotion categories based on psychological theories and validated through machine learning models. We select the five most commonly used emotion classification methods in large emotion datasets, propose a valid mapping method rooted in psychological theory, verify it through human evaluation, and assess its impact on emotion recognition models. Our mapping method is an initial effort to create a continuous mapping approach connecting these discrete emotion classification methods.

2 Methods

2.1 Datasets

We choose 4 diverse datasets, each employing distinct modalities and emotion classification methods. We include both datasets that reflect real-life scenarios such as MEmoR (Shen et al., 2020) and MELD (Poria et al., 2019), and those focusing on facial features like IEMOCAP (Busso et al., 2008). Additionally, we include the FER-2013 (Goodfellow et al., 2013) computer vision dataset to investigate our mapping method's impact on a single-modality dataset. These datasets span various classification methods: MEmoR employs Plutchik's Wheel of Emotion (14 emotions), MELD and IEMOCAP adopt Ekman's basic 6 emotions, and FER-2013 features 7 common emotions as labels.

2.2 Mapping Method

Our approach to developing a mapping theory between emotion classification methods follows the following procedure.

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. So it it possible that in the annotation process annotators sometimes use common sense understanding of emotions to annotate and only use the definitions provided as references. Given these considerations and results, we decide not to modify emotions common to both categories.

Higher-Level Emotions: Emotions exclusive to higher-level categories, such as anticipation and surprise, are mapped based on past literature, often considering valence and arousal of various emotions. Valence measures the positiveness or negativity of an emotional stimulus (De Silva et al., 1997), and emotions with similar valence are presumed to be more closely related. Arousal level, measuring the intensity of emotion, is also a cue to the similarity of emotions. 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. Detailed evaluations are carried out for each theory. We illustrate our mapping choice for the emotion "surprise" as an example of our decision-making process.

2.3 The Classification for Surprise as Example

Surprise characterizes the feeling of shock due to perceiving things or experience out of expectation. To map surprise onto a 6-emotion classification (neutral, sadness, joy, disgust, anger, and fear), we employed a bipolar model integrating valence and arousal dimensions. Russell introduced this model in 1977 (Russell and Mehrabian, 1977), with motivation as an initial component. Surprise may be considered a negative emotion, since previous studies associate surprise with a negative valence (Noordewier and Breugelmans, 2013) and high arousal levels (Russell and Mehrabian, 1977). Based on Liu et al.'s research, high-arousal, low-valence emotions are akin to anger (Liu et al., 2010). 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

<table>
<tr><th>14 fine-grained</th><th>9 primary</th><th>7 basic</th><th>6 emotions</th><th>3 sentiments</th></tr>
<tr><td>anticipation</td><td rowspan="2">anticipation</td><td rowspan="3">neutral</td><td rowspan="3">neutral</td><td rowspan="3">neutral</td></tr>
<tr><td>interest</td></tr>
<tr><td>neutral</td><td>neutral</td></tr>
<tr><td>fear</td><td>fear</td><td>fear</td><td>fear</td><td rowspan="8">negative</td></tr>
<tr><td>disgust</td><td rowspan="2">disgust</td><td rowspan="2">disgust</td><td rowspan="2">disgust</td></tr>
<tr><td>boredom</td></tr>
<tr><td>sadness</td><td>sadness</td><td>sadness</td><td>sadness</td></tr>
<tr><td>anger</td><td rowspan="2">anger</td><td rowspan="2">anger</td><td rowspan="4">anger</td></tr>
<tr><td>annoyance</td></tr>
<tr><td>surprise</td><td rowspan="2">surprise</td><td rowspan="2">surprise</td></tr>
<tr><td>distraction</td></tr>
<tr><td>joy</td><td rowspan="2">joy</td><td rowspan="3">joy</td><td rowspan="3">joy</td><td rowspan="3">positive</td></tr>
<tr><td>serenity</td></tr>
<tr><td>trust</td><td>trust</td></tr>
</table>

Table 1: Mapping results. This table 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.

mean differences (0.058) are lowest for the neutral-surprise pairing (Mancini et al., 2018). 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. All annotators were English-speaking college students, with half of them familiar with the TV show "The Big Bang Theory." Annotation materials included clips, scripts, and emotion definitions per category. Evaluation results favored the surprise-to-anger conversion, as it achieved higher accuracy. Hence, we chose to map surprise to anger based on annotation outcomes.

2.4 The Annotation Process

At least two annotators are asked to annotate one data point. All annotators are college students studying in a university where English is the first language, since the datasets are all in English. The students age between 18 to 22. The annotators are provided with clips and scripts during the annotation, and half of the annotators are familiar with the TV show, the Big Bang Theory. The emotions and definitions of each emotion in each category are also provided to the annotators to help interpretation.

3 Mapping Between Different Emotion Categories

3.1 Mapping Results

Table 1 shows the resulting unique mapping table between the 5 most popular emotion classification methods, ranging from 14 categories to 3 categories. To validate our mapping, a re-annotation of randomly sampled emotions mapped to their categories achieves an accuracy of 0.96 (Annotator 1) and 0.917 (Annotator 2), with a fair inter-annotator agreement of 0.318 (Cohen's Kappa). Thus, this mapping method has proved to have fairly high accuracy when used to reconstruct datasets. We conclude that it is possible to map emotion categories onto each other with relatively high accuracy. The proposed mapping method is one directional, from more categories to fewer categories. Mapping data from fewer categories to more categories is possible but requires additional annotation to determine the resulting co-domain labels. Additionally, this mapping method can be used by future researchers with more fine grained labeling methods when creating datasets, since mapping from more fine grained labeling to less fine grained labeling requires no additional information.

3.2 Map analysis

The main contribution of our work is that we are the first to propose a mapping method for numerous emotion categorization methods from psychological theories and have validated it with human evaluation and experiments. Analyzing the final mapping produced, we found that across all categorization methods, the categories in negative emotions are more fine-grained than either positive or neutral emotions, given the number of emotions that are mapped into negative emotions. For example, from the 14-categories, there are 8 emotions that were mapped into "negative", 3 mapped into "positive" and 3 mapped into "neutral". This imbalance could be caused by both biases in the dataset and underlying psychological mechanisms. Since the data for the datasets are collected from TV shows or other commercialized media, it could be that a dataset may not necessarily contain emotion proportions that are reflective of actual human emotional expressions. The underlying psychological mechanisms would also be an aspect to discuss for other researchers.

Moreover, while several emotions seem more difficult to be mapped into other categories, such as surprise and trust, in the experiment we found it still has an acceptable evaluation score. For example, it is difficult to determine whether 'surprise' is a good surprise or a bad surprise in real life, but in our mapping, 'surprise' is mapped into anger with a high agreement in human evaluation. One possible reason for this is that the current categories make humans, the annotators, more likely to choose negative surprise as "surprise" and consider taking positive surprise as "joy" or "hopeful". We attribute this alignment to the disparity among emotion classification theories and their unique aspects in understanding human emotions. Nevertheless, our mapping method establishes a consistent standard grounded in existing datasets.

Although the same emotion categories may have different definitions for different classification methods, each of the emotions are still mapped into the corresponding emotion with the same name in our mapping. Although we acknowledge the slight difference in meaning, for the purpose of mapping, emotions still prove to be more similar to corresponding emotion with the same name despite the different interpretations. Our current mapping method sucessfully proposes a uniform standard, yet its accuracy is limited in datasets that are largely different from the existing datasets in terms of domain, conversation style, etc. Furthermore, since we are the first to propose a mapping for different emotion classification theories from a psychological perspective, there are a limited number of existing studies that we could compare to. We hope our proposal, as a first attempt to solve this disparity, could also serve as a start point for others who seek to solve the problem.

4 Mapping effects on ML Models

To analysis the effect of the proposed mapping on machine learning models, we set up an experiment to check the accuracy of emotion re-categorization after applying the mapping method in Table 1 to both the MEmoR and the CNN dataset. We selected two models to study the effect of the mapping methods on emotion detection models.

4.1 Models

Vision CNN is commonly used in recognition and classification tasks (Albawi et al., 2017; Suryani et al., 2016). We reconstructed the FER-2013 Dataset (7 basic emotions) based on our mapping in Table 1 to recreated the dataset with 6 emotions and 3 sentiments respectively.

Multimodality MEmoR Model (Shen et al., 2020) is a fusion multi-modal model is provided by (Shen et al., 2020). The model extracts representative multimodal features, including audio features, video features, and text features, personality features, and uses an attention-based multimodal reasoning method. In the experiment we use the MEmoR dataset reconstructed based on our mapping, which has 5 groups of labels. Each model will be trained tested on each classification method.

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 2: 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 highest achieved is bolded. The MEmoR model uses visual, audio, textual features. In the CNN model, only visual information is used.

Figure 1: 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.

4.2 Results

Results of the experiments on the MEmoR model and CNN model are shown in Table 2. From these experiments, we have found that models generally perform better when there are fewer emotion categories, meaning that more fine-grained emotions are more difficult for models to differentiate, regardless of which modality or which combination of modalities is used. This finding validates that our mapping is accurate, as it is the general understanding in the machine learning community that using fewer classification categories, when correctly applied, leads to higher accuracy since the complexity of the task is reduced. However, the experimental results for the MEmoR model show that training and testing on 7 categories does achieve

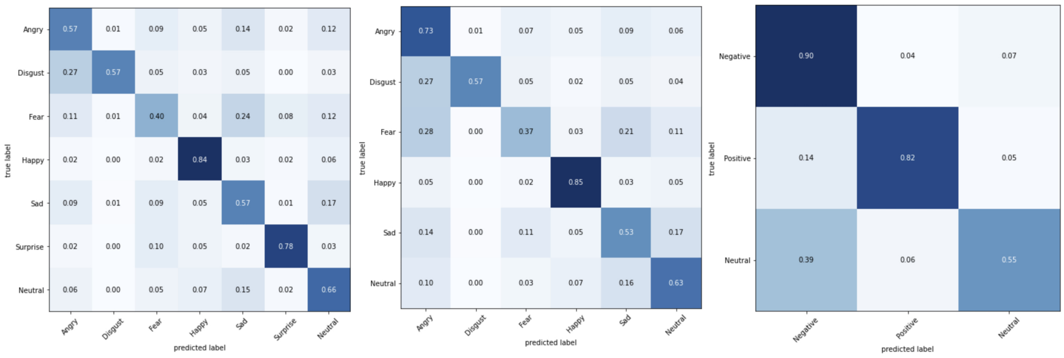

Figure 2: Confusion matrices generated by three CNN models trained on a dataset, all learning from the same set of pictures but with labels categorized into 7 (left), 6 (middle) and 3 categories (left). Columns represent the predicted label and rows represent the true label.

higher accuracy than 6 categories, while still lower than results on 3 categories. However, on the CNN model, we see a higher accuracy on 6 categories compared to the 7 categories. By looking closely at the confusion matrices (Figure 2) of CNN models, we see that the improvement was mainly on the adjusted category, and the accuracy of the categories that remain untouched from the transition remains in the same range. A possible reason for this is that classifying emotion into 7 categories is derived from Ekman's basic emotion theory, which is based on facial expression. Thus it is possible that such a categorization method is easier for models to learn through facial expression recognition. However, to determine the cause, there should be more research on separated models and modalities. We encourage future researchers to look into this question.

Visualization of the CNN model's attention is shown in Table 1. We observe that the attention of the model trained with more fine-grained emotions is more spread out through the face, with some stress around the eye and mouth area. In comparison, the attention of the model trained on sentiments is more focused on specific areas and created red dots on the heat map. The difference indicates that there are more subtle cues to distinguish fine-grained emotion on the face, requiring the model to learn to predict based on more information from different areas, compared to sentiments that are simpler and distinguishable through some key area like the mouth (smiling or not, for example).

5 Conclusion

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 are the first group of researchers attempting 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.

Acknowledgements

We thank Dr.Robert Brotherton, Bingpu Zhao, and Jingyi Song for their support and valuable feedback. This research is supported in part by the Defense Advanced Research Projects Agency (DARPA), via the The Computational Cultural Understanding (CCU) program contract HR001122C0034. The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

Limitations

A limitation of our mapping is that it proposes a unified standard within a set range of 3 to 14 categories. Yet for some particular tasks, creating a recognizer that is sensitive to a particular facial expression or emotion that is not included in our proposed method may be necessary. We encourage future researchers to expand on top of our classification method using similar methods. However, we hope that providing a unified standard would benefit the community by decreasing deviance and making it easier for scholars who wish to adopt an existing dataset for a particular task.

Moreover, while several emotions seem harder to be mapped into other categories, we found acceptable evaluation score for the mapping, but there are limitations. Similarly to the mapping of "surprise", whether the emotion “trust” was a neutral emotion or a positive emotion is hard to decide. In our classification, we followed the steps described in our “Methods” section to determine which classification gives better accuracy and thus determines the mapping. Although our current mapping method proposes a uniform standard, its accuracy is limited in datasets that are largely different from the existing datasets in terms of domain, conversation style, etc. we also acknowledge potential difficulties in mapping certain emotions, and we anticipate revisions and improvements to our current mapping method after the construction of larger datasets in the future to better bridge the differences between various data sets.

Furthermore, since we are the first to propose a mapping for different emotion classification theories from a psychological perspective, there is a limited number of existing studies that we could compare to. We hope our proposal, as a first attempt to solve this disparity, could also serve as a start point for others who seek to solve the problem.

References

Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. 2017. Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)*, pages 1–6.

Laura-Ana-Maria Bostan and Roman Klinger. 2018. An analysis of annotated corpora for emotion classification in text. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2104–2119, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower Provost, Samuel Kim, Jeannette Chang, Sungbok Lee, and Shrikanth Narayanan. 2008. Iemocap: Interactive emotional dyadic motion capture database. *Language Resources and Evaluation*, 42:335–359.

Luna De Bruyne, Orphee De Clercq, and Veronique Hoste. 2020. An emotional mess! deciding on a framework for building a Dutch emotion-annotated corpus. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 1643–1651, Marseille, France. European Language Resources Association.

L.C. De Silva, T. Miyasato, and R. Nakatsu. 1997. Facial emotion recognition using multi-modal information. In *Proceedings of ICICS, 1997 International Conference on Information, Communications and Signal Processing. Theme: Trends in Information Systems Engineering and Wireless Multimedia Communications (Cat.*, volume 1, pages 397–401 vol.1.

Paul Ekman. 1992. Are there basic emotions? *Psychological review*, 99 (3).

Ziwei Gong, Qingkai Min, and Yue Zhang. 2023. Eliciting rich positive emotions in dialogue generation. In *Proceedings of the First Workshop on Social Influence in Conversations (SICon 2023)*, pages 1–8, Toronto, Canada. Association for Computational Linguistics.

Ian J. Goodfellow, Dumitru Erhan, Pierre Luc Carrier, Aaron Courville, Mehdi Mirza, Ben Hamner, Will Cukierski, Yichuan Tang, David Thaler, Dong-Hyun Lee, Yingbo Zhou, Chetan Ramaiah, Fangxiang Feng, Ruifan Li, Xiaojie Wang, Dimitris Athanasakis, John Shawe-Taylor, Maxim Milakov, John Park, Radu Ionescu, Marius Popescu, Cristian Grozea, James Bergstra, Jingjing Xie, Lukasz Romaszko, Bing Xu, Zhang Chuang, and Yoshua Bengio. 2013. Challenges in representation learning: A report on three machine learning contests.

Ursula Hess. 2017. Chapter 5 - emotion categorization. In Henri Cohen and Claire Lefebvre, editors, *Handbook of Categorization in Cognitive Science (Second Edition)*, second edition edition, pages 107–126. Elsevier, San Diego.

Holger Hoffmann, Andreas Scheck, Timo Schuster, Steffen Walter, Kerstin Limbrecht, Harald C. Traue, and Henrik Kessler. 2012. Mapping discrete emotions into the dimensional space: An empirical approach. In *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 3316–3320.

Samira Ebrahimi Kahou, Xavier Bouthillier, Pascal Lamblin, Caglar Gulcehre, Vincent Michalski, Kishore Konda, Sébastien Jean, Pierre Froumenty, Yann Dauphin, Nicolas Boulanger-Lewandowski, Raul Chandias Ferrari, Mehdi Mirza, David Warde-Farley, Aaron Courville, Pascal Vincent, Roland Memisevic, Christopher Pal, and Yoshua Bengio. 2015. Emonets: Multimodal deep learning approaches for emotion recognition in video.

Yisi Liu, Olga Sourina, and Minh Khoa Nguyen. 2010. Real-time eeg-based human emotion recognition and visualization. In *2010 International Conference on Cyberworlds*, pages 262–269.

Giacomo Mancini, Roberta Biolcati, Sergio Agnoli, Federica Andrei, and Elena Trombini. 2018. Recognition of facial emotional expressions among italian pre-adolescents, and their affective reactions. *Frontiers in psychology*, 9:1303.

Marret K. Noordewier and Seger M. Breugelmans. 2013. On the valence of surprise. *Cognition and Emotion*, 27(7):1326–1334. PMID: 23560688.

Robert Plutchik. 2001. 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):344–350.

Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. 2019. MELD: A multimodal multi-party dataset for emotion recognition in conversations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 527–536, 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.

Guangyao Shen, Xin Wang, Xuguang Duan, Hongzhi Li, and Wenwu Zhu. 2020. Memor: A dataset for multimodal emotion reasoning in videos. In *Proceedings of the 28th ACM International Conference on Multimedia*, MM '20, page 493–502, New York, NY, USA. Association for Computing Machinery.

Dewi Suryani, Patrick Doetsch, and Hermann Ney. 2016. On the benefits of convolutional neural network combinations in offline handwriting recognition. In *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 193–198.

Christine D. Wilson-Mendenhall, Lisa Feldman Barrett, W. Kyle Simmons, and Lawrence W. Barsalou. 2011. Grounding emotion in situated conceptualization. *Neuropsychologia*, 49(5):1105–1127.

A Appendix

A.1 Experiment Design for CNN Model

To explore the effects of our mapping method on CNN models, we built a simple CNN model with three convolutional layers, feeds into a fully connected layer, and outputs from a softmax layer. The model is trained on unimodal (visual) information on the FER-2013 (Goodfellow et al., 2013) dataset for emotion classification. The CNN model is selected to study the effect of the mapping methods on unimodal models.The model was trained using batch size=256 for 60 epoches on single GPU. We reconstructed the FER-2013 (Goodfellow et al., 2013) Dataset based on our mapping. Since the dataset is originally classified labeled with 7 basic emotions, we recreated the dataset with 6 emotions and 3 sentiments classification methods respectively (Table 3 (Appendix)). The mapping method is shown in Figure 3 (Appendix). Each CNN model will be tested on all 3 classification methods using the same hyper-parameter and trained for 60 epochs in two stages on the same hardware. All three models are trained to convergence before stopping at epochs 60.

A.2 Experiment Design for MEmoR Model

MEmoR Model (Shen et al., 2020) is a fusion multi-modal model is provided by (Shen et al., 2020). The model extracts representative multimodal features, including audio features, video features, and text features, personality features, and uses an attention-based multimodal reasoning method. The experiment use the MEmoR (Shen et al., 2020) dataset reconstructed based on our mapping. The reconstructed dataset has 5 groups of labels, following the 5 most popular emotion classification theories. Each model will be tested on all 5 classification methods and each modality (visual, textual, audio) in order to explore the effect of our mapping on models. For simplicity, we choose the default parameters and model structure given in the MEmoR model, except to revise the model to fit the change in the size of the label. All 5 classification methods experimented with are listed in Table 3 (Appendix). The mapping method is shown in Figure 3 (Appendix).

14 fine-grained emotions	9 primary emotions	7 basic emotions	6 emotions	3 sentiments
joy, anger, disgust, sadness, surprise, fear, anticipation, trust, serenity, interest, annoyance, boredom, distraction, neutral	joy, anger, disgust, sadness, surprise, fear, anticipation, trust, neutral	joy, anger, disgust, sadness, fear, surprise, neutral	joy, anger, disgust, sadness, fear, neutral	positive, negative neutral

Table 3: Emotion Categories

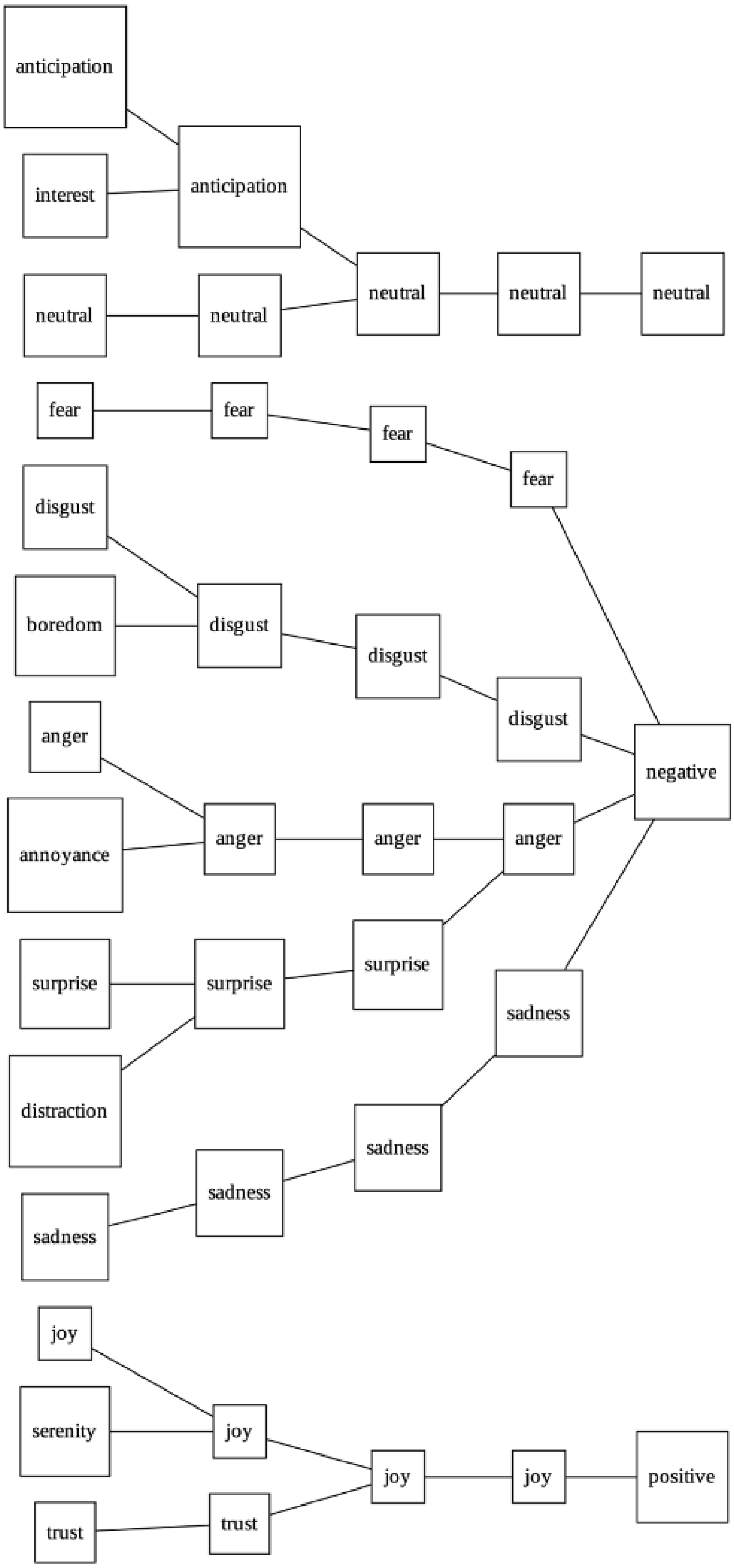

Figure 3: Mapping method in graph. This graph demonstrates how 14 fine-grained emotions, listed on the leftmost column, are mapped onto 9 primary emotions, Ekman's basic emotions, 6 emotions, and the 3 sentiments.