This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2025.3545447

Learning Emotion Category Representation to Detect Emotion Relations across Languages

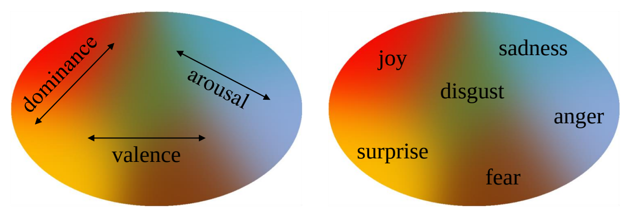
Xiangyu Wang, and Chengqing Zong, *IEEE Fellow*

Abstract—
Understanding human emotions is crucial for a myriad of applications, from psychological research to advancements in Natural Language Processing (NLP). Traditionally, emotions are categorized into distinct basic groups, which has led to the development of various emotion detection tasks within NLP. However, these tasks typically rely on one-hot vectors to represent emotions, a method that fails to capture the relations between different emotion categories. In this study, we challenge the assumption that emotion categories are mutually exclusive and argue that the connections and boundaries between them are complex and often blurred. To better represent these nuanced interconnections, we introduce an innovative framework as well as two algorithms to learn distributed representations of emotion categories by leveraging soft labels from trained neural network models. For the first time, our approach enables the detection of emotion relations across different languages through an NLP lens, a feat unattainable with traditional one-hot representations. Validation experiments confirm the superior ability of our distributed representation algorithms to articulate these emotional connections. Moreover, application experiments corroborate several interdisciplinary insights into cross-linguistic emotion relations, findings that align with research in psychology and linguistics. This work not only presents a breakthrough in emotion detection but also bridges the gap between computational models and humanistic understanding of emotions.

Index Terms—Emotion category, emotion space, distributed representation, emotion relations across languages.

1 Introduction

EMOTIONS are subjective experiences that people represent with language [1]. People express their emotions to others with various communication methods such as facial expressions, speech, text, body language, etc. As shown in Figure 1, there are two main approaches to describe the distribution of human emotional states: predefined dimensional approach and namely categorical approach [2]. In the predefined dimensional approach, emotions are quantitatively represented with several predefined attributes. The most popular framework of this kind was proposed by Russell, considering *valence* (positiveness–negativeness) and *arousal* (active–passive) as the two core attributes of all emotional experiences [3]. Mehrabian and Russell introduced three dimensions to represent the emotional state: *pleasure* (pleasure–displeasure), *arousal* and *dominance* (dominant–submissive), which is known as the PAD model [4]. Nevertheless, on the one hand, two or three attributes are not enough to describe the wide range of human emotional states [5], [6]. On the other hand, it is quite difficult to quantify the attribute value of a specific emotional state [7]. In namely categorical approach, human emotional states are presented with discrete basic emotion categories. The most well-known model of this kind is proposed by Ekman, who introduced the existence of six basic emotion categories: *anger*, *disgust*, *fear*, *joy*, *sadness* and *surprise* [8]. Cowen and Keltner introduced a conceptual framework to analyze reported emotional states and elicited 27 distinct varieties of reported emotional experience [9]. Many other basic emotion theories provide different clusters, ranging from 6 to 15 [10], [11], [12]. Nonetheless, on the one hand, the categorical models divide human emotional states into limited emotion categories. Thus, the different emotional states may correspond to the same emotion category. On the other hand, the categorical models treat emotion categories as independent ones, which ignore the underlying relations between emotion categories.



(a) Dimension Model (b) Categorical Model

Fig. 1. An illustration of two main models to describe human emotional states, which is represented by the multi-color ellipse.

For an emotional state contained in a document, it is much easier to label its emotion category than to annotate the value of its specific attributes. As a result, the basic emotion categories have been widely applied in text emotion analysis in the field of natural language processing (NLP). Many downstream tasks and corresponding datasets have been proposed in the past decades. To recognize the emotional content contained in the text, Alm et al. introduced machine learning methods to predict emotion

- *The authors are with the State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences and University of Chinese Academy of Sciences, No. 95 Zhongguancun East Road, Beijing 100190, P.R. China, and also with the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, P.R. China.*
- *Xiangyu Wang is now with the E Fund Management Co., Ltd, Guangzhou 510620, P.R. China.*
- *E-mail: {xiangyu.wang, cqzong}@nlpr.ia.ac.cn*

Manuscript received May 14, 2024.
(Corresponding author: Chengqing Zong)

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on March 19,2025 at 08:46:11 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2025.3545447

categories from the text in the domain of children's fairy tales for the first time [13]. Lee et al. first proposed a linguistic-driven rule-based system for emotion cause detection, as well as constructed a Chinese emotion cause corpus annotated with emotions and the corresponding cause events [14]. Thet et al. found the sentiment towards various aspects in a document (such as cast, director, story, and music in a movie review) may be different, and they employed linguistic methods to study aspect-based sentiment analysis in the field of the movie review [15]. Jiang et al. proposed a target-dependent sentiment classification task on Twitter, in which the tweets are supposed to be classified as positive, negative, or neutral according to the given query [16]. Zhou et al. first introduced the emotion distribution learning task to identify multiple emotions with their intensities from texts since multiple emotions with different intensities are often co-existed in a single sentence [17]. Mohammad and Bravo-Marquez first formulated the task of detecting emotion intensities in tweets [18]. Alejo et al. extend the emotion intensity prediction task to multilingual [19]. Based on the emotion cause extraction task, Xia et al. proposed a new task named emotion-cause pair extraction to solve the shortcomings of the traditional emotion cause extraction task that depends on the annotation of emotion before extracting cause [20]. There is also a lot of research in the field of multilingual emotion analysis. Abdalla et al. employ a single linear transformation to capture fine-grained sentiment relationships between cross-lingual words [21]. Dufter et al. present a method to word embedding space by concept induction [22]. Zhao and Schütze [23] propose a universal approach for sentiment lexicon induction and conduct experiments on a parallel corpus of 1593 languages.

In existing tasks and models mentioned above, emotional states in the text are described with one or several independent emotion categories, which are represented with one-hot vectors. However, the underlying relations among the emotion categories are ignored in one-hot representation, which is contrary to the fact that the emotion categories are not orthogonal to each other and the boundaries as well as emotion relations are not clearly distinguished and defined [24], [25], [26]. Therefore, the existing one-hot representations for emotion categories are subject to certain restrictions in many practical applications. For example, there is a warm debate about whether emotion categories are universal or language-specific in cognitive psychology [27], [28], [29]. Some studies indicated that human emotion categories are universal [30], while others suggested they are language-specific [31]. Since different categories are orthogonal to each other in one-hot representation, one-hot vectors cannot be employed to express emotion relations, let alone to analyze emotion relations across languages.

Different from previous studies that treated emotion categories as independent ones, we propose a novel framework to learn the distributed representations for emotion categories in this paper, which regards the collection of human emotional states contained in text as an emotion space, and each emotional state contained in a document corresponds to a point in the space. As shown in Figure 1 (b), each emotion category is a cluster distributed in the emotion space rather than a specific constant point. And the representation of the cluster center are learned as the representation of the corresponding emotion category. Based on the learned distributed representations of emotion categories, emotion relations across languages can be detected from the perspective of NLP, which cannot be achieved by existing one-hot representations.

Based on our previous conference paper [32], the following aspects have been extended in this work.

- We propose a general framework to learn distributed representations for emotion categories. Forced Symmetry mechanism is introduced to refine the emotion category representations, ensuring improved stability of emotion vectors in the emotion space.
- We conduct more experiments to validate the effectiveness of the learned emotion category representations. The results demonstrate that the learned representations of emotion categories in emotion space can express emotion relations much better than word vectors, and is competitive with human results.
- We apply the learned representations of emotion categories to analyze the differences in emotion relations across languages. As far as we know, this is the first work to study the differences of emotion relations across languages from the perspective of NLP.

The remainder of this paper is organized as follows. The related work is introduced in Section 2. Section 3 describes the limitations of existing representations and the algorithms to obtain our distributed representations of emotion categories. Section 4 reports the validation of the effectiveness of our algorithms. Section 5 illustrates the application of our methods in detecting emotion relations across languages. Section 6 summarizes the major conclusions.

2 Related Work

Four lines of related literature will be reviewed in this section: basic emotion models, emotion datasets, emotion across languages, and soft labels.

2.1 Basic Emotion Models

There are many studies on emotion taxonomy that divide human emotional states into different basic emotion categories. Weiner and Graham simply divided emotions into *happiness* and *sadness*, which is the same as using polarity to describe the emotional state [33]. Plutchik proposed an emotion wheel based on general psychoevolutionary theory, in which the human emotional states are divided into eight main categories: *anticipation*, *disgust*, *hate*, *joy*, *love*, *sadness*, *surprise*, and *trust* [34]. Ekman suggested that there are six discrete basic emotion categories (*anger*, *disgust*, *fear*, *joy*, *sadness*, and *surprise*) [8], which is the most popular emotion taxonomy by far and is contained in the vast majority of the existing emotion classification datasets. Parrott proposed another six basic types of emotion category from the perspective of social psychology (*anger*, *fear*, *joy*, *love*, *sadness*, and *surprise*) [35]. Harmon et al. captured eight distinct state emotions in their study with the questionnaire approach (*anger*, *disgust*, *fear*, *anxiety*, *sadness*, *happiness*, *relaxation*, and *desire*) [12]. Similarly, Cowen et al. introduced a conceptual framework to analyze reported emotional states and elicited 27 distinct varieties of reported emotional experience [9]. However, the basic emotion categories mentioned above are derived on the basis of psychological research on human emotional states. The relations among these emotions are still unclear. In this work, we propose a framework to learn the distributed representations of emotion categories, and the similarities among basic emotions can be further detected.

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on March 19,2025 at 08:46:11 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2025.3545447

TABLE 1
Differences between semantic space and emotion space.

Semantic Space	**Emotion Space**
Each word corresponds to a point in semantic space.	Words cannot be represented in emotion space.
Emotional states cannot be represented in semantic space.	Each emotional state corresponds to a point in emotion space.
Each emotion category is encoded with a piece of specific semantic information.	Each emotion category is encoded with a specific emotional state.

2.2 Emotion Datasets

In order to achieve a better performance on existing emotion detection tasks, many datasets that vary in the domain, language, size, and taxonomy have been introduced. A standard approach to create an emotion recognotion dataset is via *expert annotation*. Strapparava and Mihalcea first introduced an emotion dataset, Affective Text, in the field of news headlines [36]. For the purpose of improving the size of the emotion dataset, Wang et al. created a large emotion dataset automatically from about 2.5 million tweets by harnessing emotion-related hashtags available [37]. Abdul-Mageed and Ungar developed a dataset contained 24 fine-grained types of emotions from Twitter [38]. In the domain of daily dialogue, Li et al. introduced a multi-turn dialog dataset (DailyDialog) by crawling the raw data from various websites [39]. To create a multilingual emotion dataset, Öhman et al. presented a dataset in the field of movie subtitles with a gamified framework [40]. Different from annotating datasets via *expert annotation*, Scherer and Wallbottcreated ISEAR dataset with another approach named *self-reporting*, in which subjects are asked to describe situations associated with a specific emotion [41]. With *self-reporting* approach, Demszky et al. introduced GoEmotions, a very manually annotated dataset of 58k English Reddit comments and annotated with 27 emotion categories [42]. Nevertheless, the instances in existing emotion recognition datasets are annotated with one or several basic emotion categories, and the emotion categories are represented with one-hot vectors. Emotions are orthogonal to each other in one-hot representation, which does not accord with the fact that the emotion relations are very complex in real world. In this paper, we learn the distributed representations for emotion categories in emotion space, and emotion relations are further detected based on our emotion representations.

2.3 Emotions across Languages

In the field of cognitive psychology, there is a question about whether emotion is universal or language-specific. As in many other debates, there is a continuum of positions about how language dominated emotions. On the one hand, universalists argued that basic emotion categories are experienced and expressed independently of language. On the other hand, relativists thought that language affects the individual's emotions. Izard and Buechler proposed a theory of emotions that considered a fundamental emotion as a complex motivational phenomenon [43]. They described ten basic discrete emotions with corresponding universal facial expressions. Ekman et al. presented their evidence of agreement across languages in the judgment of facial expression by asking the members to show their facial expressions in different emotional contexts [44]. Matsumoto discussed how emotions are influenced by culture in detail from a historical perspective [45]. Kotz and Paulmann put forward a view that speech emotion and language comprehension are anchored in a functionally differentiated brain network from a brain science perspective [46]. Sundararajan provided an explanatory framework to cast the East and West difference in facial expression of emotions from a cognitive psychology perspective [47]. However, the above studies analyzed emotion spaces across languages from the perspective of psychology or brain science. There is still no relevant work in the NLP field as emotion categories are usually represented with one-hot vectors in existing emotion detection tasks. In this paper, based on the distributed representations of emotion categories, we study the text emotion spaces across languages from the perspective of NLP for the first time.

2.4 Soft Labels

Hinton et al. in their novel work first discussed that soft targets from a well-trained large model have more entropy and can provide more information than manually annotated hard targets [48]. Phuong and Lampert provided their insights into the working mechanisms of distillation by studying the special case of linear and deep linear classifiers [49]. Zhang et al. proposed a method to represent labels with the average word embedding of the word terms of the label name [50]. Nevertheless, label representation in semantic space cannot reflect label relations well. In this work, we represent labels in label space (i.e. emotion space) rather than semantic space, and the experimental results demonstrate that our representation method can express label relations much better than existing representation methods in semantic space. For the purpose of making the model to be less confident in the image classification task, Szegedy et al. proposed a label smoothing technique, in which the one-hot labels are regularized to a weighted mixture of targets in the dataset [51]. In order to better understand label smoothing in the context of neural machine translation, Gao et al. derived and explained theoretically why label smoothing is optimizing [52]. Nevertheless, the mixture targets are obtained only by adding a noise vector that the value in each dimension is the same. As a result, the mixture targets from label smoothing approach cannot reflect the true label distribution. In this work, soft labels output by trained neural network models are employed to generate the distributed representations for emotion categories. Furthermore, we generate a more appropriate label distribution to enhance classification tasks.

3 Motivation

3.1 Differences between Semantic Space and Emotion Space

The introduction of word vectors [53] has enabled the distributed representation of words in a high-dimensional semantic space. Each word, including emotion category terms, is represented as a vector in this space. However, word vectors are computed based on the semantic averaging hypothesis, where the vector for each emotion term is derived as the mean value of its contextual word

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on March 19,2025 at 08:46:11 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2025.3545447

TABLE 2
Four instances in dataset AffectiveText. Instance 1 and 2 are annotated with same emotion category but different valence. Instance 3 and 4 are annotated with same valence but different emotion category.

Index	Instances	Emotion	Valence
1	Goal delight for Sheva	joy	87
2	Making peace from victory over poverty	joy	39
3	New Indonesia Calamity, a Mud Bath, Is Man-Made	anger	-59
4	Waste plant fire forces 5,000 to evacuate	sadness	-59

vectors. Consequently, these vectors capture semantic or contextual information, but fail to represent specific emotional states within the emotion space. To incorporate sentiment information into word embeddings, several sentiment-enriched methods have been developed. Tang et al. introduced a learning algorithm called sentiment-specific word embedding (SSWE) [54], while Agrawal et al. proposed emotion-enriched word embedding (EWE) [55]. Despite these advancements, both general and sentiment-enriched word vectors are derived within the semantic space framework, rather than the emotion space. Importantly, emotion space is neither equivalent to semantic space nor a subspace of it. A detailed comparison between the two is presented in Table 1.

3.2 Limitations of One-hot Representation and Label Smoothing

In existing emotion detection tasks [14], [18], [20] and following datasets [36], [39], [42], emotion categories are represented with one-hot vectors. Nevertheless, in one-hot representation, each emotion category is regarded as an independent dimension. As a result, each emotion category is orthogonal to all other emotion categories, and the cosine similarity between different categories is 0.

$$\boldsymbol{V}_i^{one-hot} \cdot \boldsymbol{V}_j^{one-hot} = 0 \tag{1}$$

where i, j denote different emotion categories, and $\boldsymbol{V}_i^{one-hot}$, $\boldsymbol{V}_i^{one-hot}$ denote corresponding one-hot vectors. However, as discussed in section 3.1, the relations among emotion categories are complex. Obviously, one-hot representation ignores the underlying relations among emotion categories in emotion space.

Similarly, in label smoothing representation, all emotion categories have similar representations. The cosine similarity values between different emotion categories are equal to a constant. Although this makes different emotional categories no longer orthogonal, it still cannot distinguish the similar relationships between emotions. For example, emotion *sadness* have the same cosine similarities with emotion *remorse* and emotion *joy* in label smoothing representation.

4 Methodology

4.1 The General Framework

In contrast to one-hot representation, our motivation is to use distributed vectors to represent emotion categories. As shown in Figure 1 in section 1, namely categorical approach employs several basic emotion categories to represent human emotion states. As a result, each emotion category corresponds to a wide range of emotional states in emotion space. Therefore, learning the distributed representation of an emotion category is actually learning a vector to represent the range of the corresponding emotion category. In other words, when we learn the distributed representation of an emotion category, we learn its cluster center.

This can be further confirmed in emotion classification datasets. Take dataset SemEval-2007 task 14 [36] as an example, four typical instances are annotated with both emotion category and valence value as shown in Table 2. Instance 1 and instance 2 are annotated with the same emotion category although they are labeled with different valence values. Actually, the emotional state in instance 1 is different from that in instance 2. Instance 1 seems to be more *excited* while the emotional state in instance 2 seems to be more *hopeful*. As for instance 3 and instance 4, they are annotated with the same valence value but different emotion categories.

The above examples demonstrate that each emotion category corresponds to a wide range of emotional states. Two documents annotated with the same emotion category may have different emotional states. In this article, we regard the whole text emotional states as a vector space. Each document corresponds to a specific emotional state, and further corresponds to a specific point in the space. Furthermore, each emotion category is a cluster rather than a specific constant point in the space, which means the emotion category is a random variable distributed in the vector space.

For category K, we define x as the sample annotated with category K and $\boldsymbol{V}_K$ as the distributed representation of category K. Let $\boldsymbol{\mathcal{V}}(x)$ be the distributed representation of sample x and $p(x)$ be the probability density of sample x. Let Ω be the integral domain of x. We further use $\mathcal{L}(\boldsymbol{V}_K, \boldsymbol{\mathcal{V}}(x))$ as the distance function between $\boldsymbol{V}_K$ and $\boldsymbol{\mathcal{V}}(x)$. In order to obtain a better distributed representation for category K, we must minimize the expectation of $\mathcal{L}$. Thus, we obtain the calculation formula for specific distributed representation of category K as the following:

$$\boldsymbol{V}_K = \arg\min_{\boldsymbol{V}} \int_{\Omega} \mathcal{L}(\boldsymbol{V}, \boldsymbol{\mathcal{V}}(x)) p(x) dx. \tag{2}$$

4.2 Two Solutions to the Framework

It is difficult to obtain the specific distribution of emotion categories in emotion space. Therefore, it is impossible to calculate the distributed representations of emotion categories directly from their distribution in emotion space. Fortunately, there are many existing available emotion classification datasets, in which each instance can be regarded as a sample of the corresponding emotion category. As a comprise solution, we employ the samples in the emotion classification dataset to estimate the emotion distribution. Thus, we can rewrite Equation 2 as:

$$\boldsymbol{V}_K = \arg\min_{\boldsymbol{V}} \sum_{x \in S_K} \mathcal{L}(\boldsymbol{V}, \boldsymbol{\mathcal{V}}(x)), \tag{3}$$

where S_K is the set of all instances labeled with category K in dataset $\mathcal{D}$.

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on March 19,2025 at 08:46:11 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2025.3545447

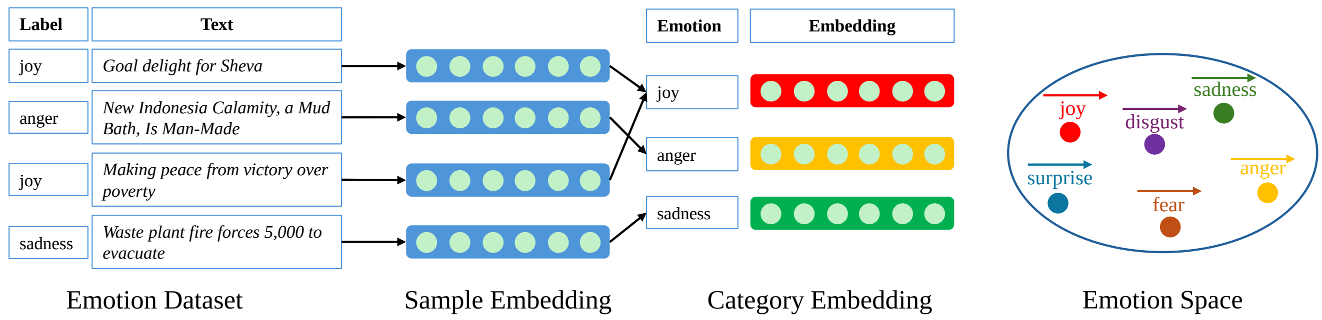


Fig. 2. The detailed steps of our approach to learn distributed representations for emotion categories. DR represents for distributed representations. Blue rectangles represent instances in the emotion dataset. Each instance in the dataset is regarded as a sample of the annotated emotion category.

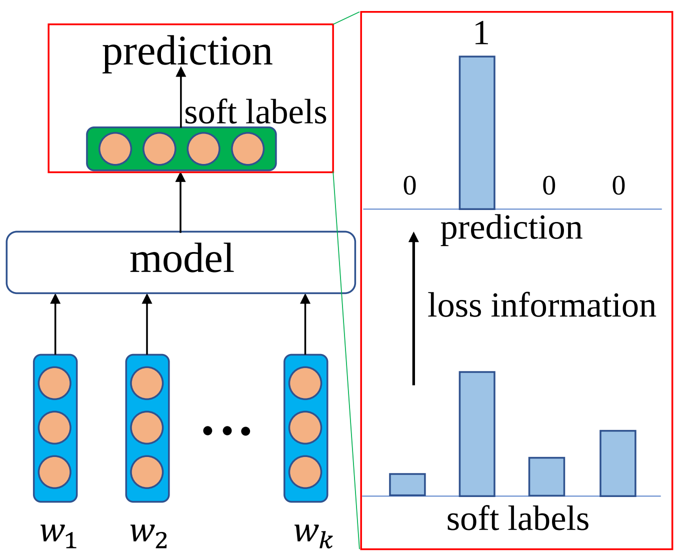


Fig. 3. Schematic diagram of neural network models for emotion classification task. For each document, the supposed input of the model is the corresponding vectors of all words in the document. The model will output the predictions based on the soft labels. The dimension of the soft labels is same to the number of categories. In this paper, temperature scaling is performed to obtain calibrated soft labels with suitable confidence levels.

The detailed approach of our method can be seen in Figure 2. In the first step, we learn the distributed representation for each instance annotated in the dataset. In the second step, we derive the distributed representation of each emotion category from the distributed representations of the instances. After completing the above two steps, we can reconstruct a high-dimensional vector space to describe human emotional states based on the distributed representations of emotion categories.

For the first step, well-calibrated soft labels output by a trained neural network model are employed directly as the distributed representation for the input instance. As shown in Figure 3, a neural network model will output soft labels regardless of its specific architecture. Previous studies [48], [49] have also verified that soft labels tend to have higher entropy and contain more information than manual one-hot labels. Since modern neural networks tend to be over-confident, *temperature scaling* [56] is employed to enable the soft labels to have a suitable confidence levels. We denote x as the input instance and $\boldsymbol{f}(x)$ as the soft labels from a trained neural network model. Thus, we have:

$$\boldsymbol{f}(x) = \boldsymbol{\mathcal{V}}(x). \tag{4}$$

For the second step, it is a classic optimization problem. Take Equation 4 into Equation 3, we have:

$$\boldsymbol{V}_K = \arg\min_{\boldsymbol{V}} \sum_{x \in S_K} \mathcal{L}(\boldsymbol{V}, \boldsymbol{f}(x)), \tag{5}$$

Depending on the choice of the loss function, the solution to this problem is not unique. In this paper, we employ squared Euclidean distance as the loss function. Therefore, Equation 5 can be simplified as follows:

$$\boldsymbol{V}_K = \arg\min_{\boldsymbol{V}} \sum_{x \in S_K} ||\boldsymbol{V} - \boldsymbol{f}(x)||_2^2. \tag{6}$$

By solving Equation 6, we have:

$$\boldsymbol{V}_K = \frac{\sum_{x \in S_K} \boldsymbol{f}(x)}{N_K}, \tag{7}$$

where N_K is the size of S_K.

Repeating the above approach for all emotion categories, we can obtain the representation matrix of emotion categories based on soft labels.

$$\boldsymbol{V}^{SL} = [\boldsymbol{V}_1; \boldsymbol{V}_2; ...; \boldsymbol{V}_C], \tag{8}$$

where SL stands for soft labels, the i-th row in $\boldsymbol{V}^{SL}$ is the distributed representation for i-th emotion category.

It should be noted that the dimension of each emotion category is equal to the dimension of soft labels, which is further equal to the numbers of emotion categories contained in the dataset. Therefore, $\boldsymbol{V}^{SL}$ is a square matrix.

Considering the K-th row in V^{SL}, which is also the representation of K-th emotion category, we have the formula:

$$\boldsymbol{V}_K = [V_{K1}; V_{K2}; ...; V_{KC}]. \tag{9}$$

The value of V_{Ki} is positively correlated with the similarity of K-th and i-th emotion category. Therefore, we suppose that a symmetric emotion category representation matrix should be better. Based on the consideration of symmetry, we force the emotion category representation matrix V^{SL} to be symmetrical.

Forced Symmetry leverages the inherent symmetry in emotional category relationships to better align distributed representations with psychological theories. This ensures that the learned emotion space more accurately reflects the natural structure of

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on March 19,2025 at 08:46:11 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2025.3545447

TABLE 3
Emotion categories in dataset GoEmotions. There are 27 emotion categories in GoEmotions, which are divided into three parts (positive, negative and ambiguous) by the authors of GoEmotions.

Positive(P):	admiration, amusement, approval, caring, desire, excitement, gratitude, joy, love, optimism, pride, relief
Negative(N):	anger, annoyance, disappointment, disapproval, disgust, embarrassment, fear, grief, nervousness, remorse, sadness
Ambiguous(A):	confusion, curiosity, realization, surprise

emotions. From a numerical stability perspective, Forced Symmetry also reduces the sensitivity of representations to random initialization and training noise by imposing constraints on the similarity matrix. This leads to more consistent and robust emotion vectors. After adopting forced symmetry to the solution $\boldsymbol{V}^{SL}$, we can obtain another solution of emotion representation matrix.

$$\boldsymbol{V}^{FS} = (\boldsymbol{V}^{SL} + \boldsymbol{V}^{SL}.T)/2, \tag{10}$$

where FS is the abbreviation for forced symmetry, $\boldsymbol{V}^{SL}.T$ represents the transpose of matrix $\boldsymbol{V}^{SL}$.

4.3 Extension to Multilabel Datasets

The above derivation to calculate the distributed representations for emotion categories is based on the single-label dataset. However, there are many datasets in which the instances are annotated with multiple emotion categories [36], [42]. In this part, we extend our approach to multilabel datasets.

We think the weight of each instance should be the same as one unit, no matter it is single labeled or not. As for multilabel instances, they can be divided into multiple single-label instances. However, the weight of each class in multilabel instances remains unknown. Based on the principle of maximum entropy [57], all classes should have the same weight. The weight of each single label data is set to the reciprocal of the number of the annotated labels. For example, suppose document D is labeled with category A and B. We regard D as two half instances, one half is labeled with category A, and the other half is labeled with category B.

Let $\mathcal{Y}(x)$ denote the set of the annotated labels of sample x and $|\mathcal{Y}(x)|$ denote the size of set $\mathcal{Y}(x)$. Take above document D as an example, then $\mathcal{Y}(D)$ is equal to $\{A, B\}$ and $|\mathcal{Y}(D)|$ is equal to 2 as there are two labels contained in $\mathcal{Y}(D)$. Therefore, we obtain the calculation formula of specific distributed representation for category K:

$$\boldsymbol{V}_K = \frac{\sum_{x \in S_K} w_K(x) \boldsymbol{f}(x)}{\sum_{x \in S_K} w_K(x)}, \tag{11}$$

where $w_K(x)$ is equal to $1/|\mathcal{Y}(x)|$, which is the weight of instance x in category K.

4.4 The Algorithm

As $\boldsymbol{V}^{SL}$ is the premise of calculating $\boldsymbol{V}^{FS}$, we propose only one algorithm to obtain $\boldsymbol{V}^{SL}$ and $\boldsymbol{V}^{FS}$ at the same time. The pseudocode of the algorithm to learn the **D**istributed **R**epresentations of **E**motion **C**ategories (DREC) is listed in Algorithm 1.

Algorithm 1 The Algorithm of DREC-SL and DREC-FS

Input: $\mathcal{D} = \{(\mathcal{T}^{(n)}, \mathcal{Y}^{(n)})_{n=1}^{N}\}$ // dataset
Output: $\boldsymbol{V}^{SL} = \{\boldsymbol{V}_1^{SL}, \boldsymbol{V}_2^{SL}, ..., \boldsymbol{V}_C^{SL}\}$
$\boldsymbol{V}^{FS} = \{\boldsymbol{V}_1^{FS}, \boldsymbol{V}_2^{FS}, ..., \boldsymbol{V}_C^{FS}\}$
// distributed representations for emotions
$\boldsymbol{f} \leftarrow \mathcal{D}$ // train a neural network model
$\boldsymbol{V}^{SL} \leftarrow \{\boldsymbol{0}, \boldsymbol{0}, ..., \boldsymbol{0}\}$
$\{W_1, W_2, ..., W_C\} \leftarrow \{0, 0, ..., 0\}$ // weight
for $n = 1$ to N **do**
  for each $j \in \mathcal{Y}^{(n)}$ **do**
    $\boldsymbol{SL} \leftarrow \boldsymbol{f}(\mathcal{T}^{(n)})$ // soft labels
    $\boldsymbol{V}_j^{SL} \leftarrow \boldsymbol{V}_j^{SL} + \boldsymbol{SL}/|\mathcal{Y}^{(n)}|$
    $W_j \leftarrow W_j + 1/|\mathcal{Y}^{(n)}|$
  end for
end for
for $i = 1$ to C **do**
  $\boldsymbol{V}_i^{SL} \leftarrow \boldsymbol{V}_i^{SL}/W_i$
end for
$\boldsymbol{V}^{FS} \leftarrow (\boldsymbol{V}^{SL} + \boldsymbol{V}^{SL}.T)/2$

5 VALIDATION

In this section, we focus on the validation of our approaches in expressing emotion relations. We first introduce the dataset and the models and then conduct three experiment to verify the effectiveness of our approaches.

5.1 Datasets

GoEmotions [42] is chosen as the dataset to validate the intrinsic quality of our approaches as there are 27 emotion categories in GoEmotions. GoEmotions is annotated of 58k English Reddit comments, with comments extracted from popular English subreddits. In contrast to Ekman's taxonomy, which includes only one positive emotion (*joy*), GoEmotions is created for the purpose of building a large dataset with a large number of positive, negative, and ambiguous emotion categories. As a result, GoEmotions can better reflect the complex relations among emotion categories. The detailed emotion categories are shown in Table 3.

5.2 Models

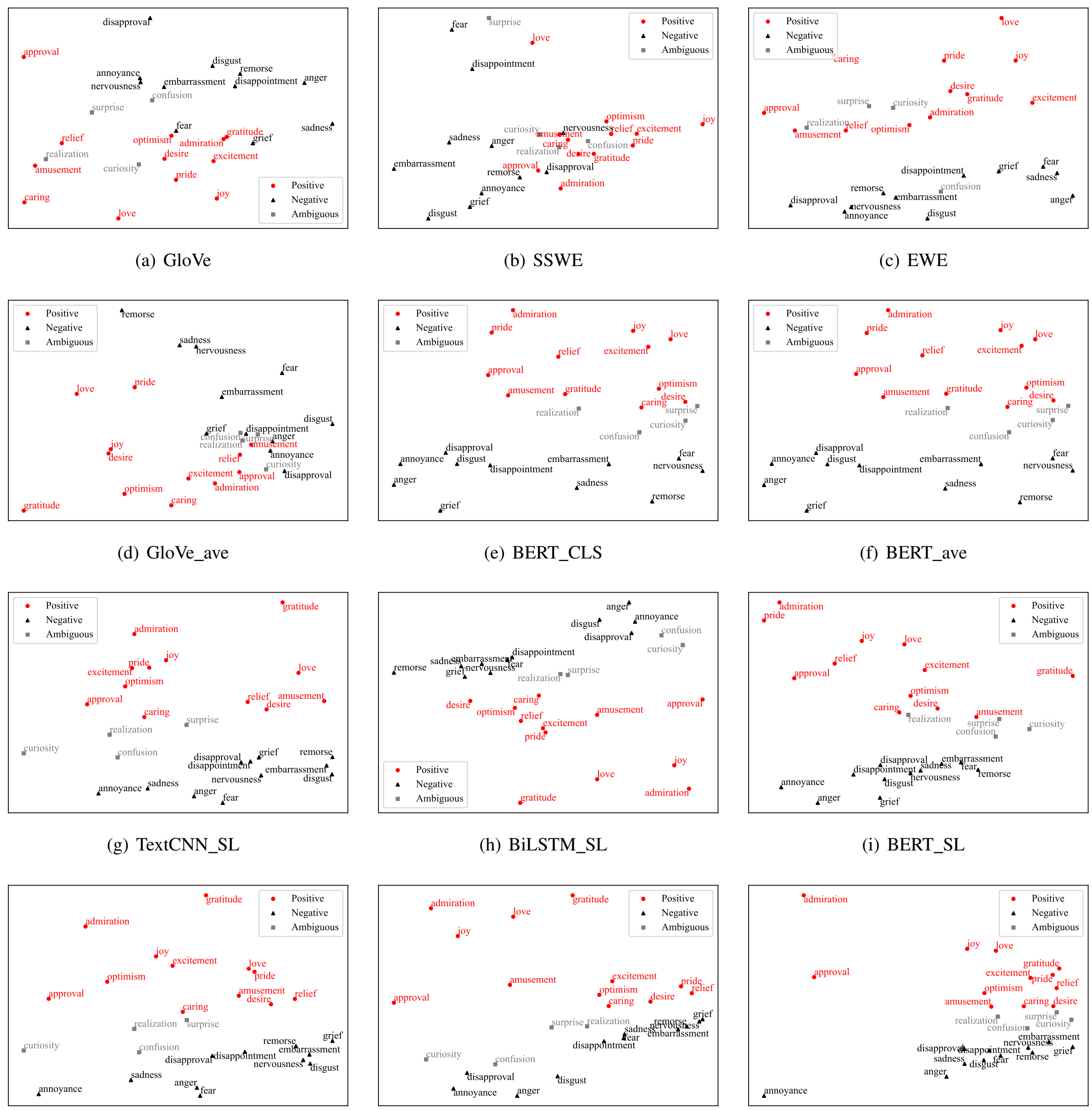
As mentioned before, any neural network model can be applied to learn distributed representations for emotion categories with our methods. Three typical neural network model, TextCNN [58], BiLSTM [59] and BERT [60], are employed in this section. The experimental results in semantic space are also conducted for comparison. 300-dimensional GloVe [61] of the term of emotion categories are chosen as its word embedding representation in semantic space. The detailed model settings are listed as follows:

TextCNN: The height of convolution kernel size is divided into three groups 2,3,4,5 and 300-dimensional random word vectors are employed. There are 32 channels in each group. Batch size and learning rate are set to 128 and 0.001.

BiLSTM: There is only one layer in this model. Batch size and learning rate are set to 128 and 0.001 separately. 300-dimensional

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on March 19,2025 at 08:46:11 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2025.3545447



(a) GloVe (b) SSWE (c) EWE

(d) GloVe_ave (e) BERT_CLS (f) BERT_ave

(g) TextCNN_SL (h) BiLSTM_SL (i) BERT_SL

(j) TextCNN_FS (k) BiLSTM_FS (l) BERT_FS

Fig. 4. Visualization of emotion vectors in different spaces. (a)-(f) In (emotion-enriched) semantic space, the linear boundaries among positive, negative and ambiguous emotion categories only exist in BERT_CLS and BERT_ave. (g)-(l) In emotion space, regardless of the choice of neural network models or algorithms, each type of emotion categories is linear separated with the others. The ambiguous emotion categories are just located between positive and negative emotion categories. (SL: soft labels; FS: forced symmetry.)

random word vectors are employed.. There are 32 neurons in the hidden layer in each direction.

BERT: BERT-based model is used in this experiment [60]. A fully connected layer is added on top of the pre-trained model. Batch size and learning rate are separately set to 128 and 2e-5 for fine-tuning.

Besides, three more baselines (GloVe_ave, BERT_CLS, and BERT_ave) that leverage GoEmotions dataset with our proposed Equation 11 are conducted to construct the emotion-enriched semantic space. For GloVe_ave (and BERT_ave), the averaged context embedding from GloVe (and fine-tuned BERT model) is chosen as the instance representation. For BERT_CLS, the embedding of $[CLS]$ token from fine-tuned BERT model is chosen as the instance representation.

5.3 Arrangement

This section presents the arrangement of emotion representations in the emotion space, alongside a comparison with the arrangement of emotion category word vectors in semantic space. The GoEmotions dataset includes 27 annotated emotion categories, detailed in Table 3. As described by its creators [42], these categories are grouped into three classes: positive, negative, and

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on March 19,2025 at 08:46:11 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2025.3545447

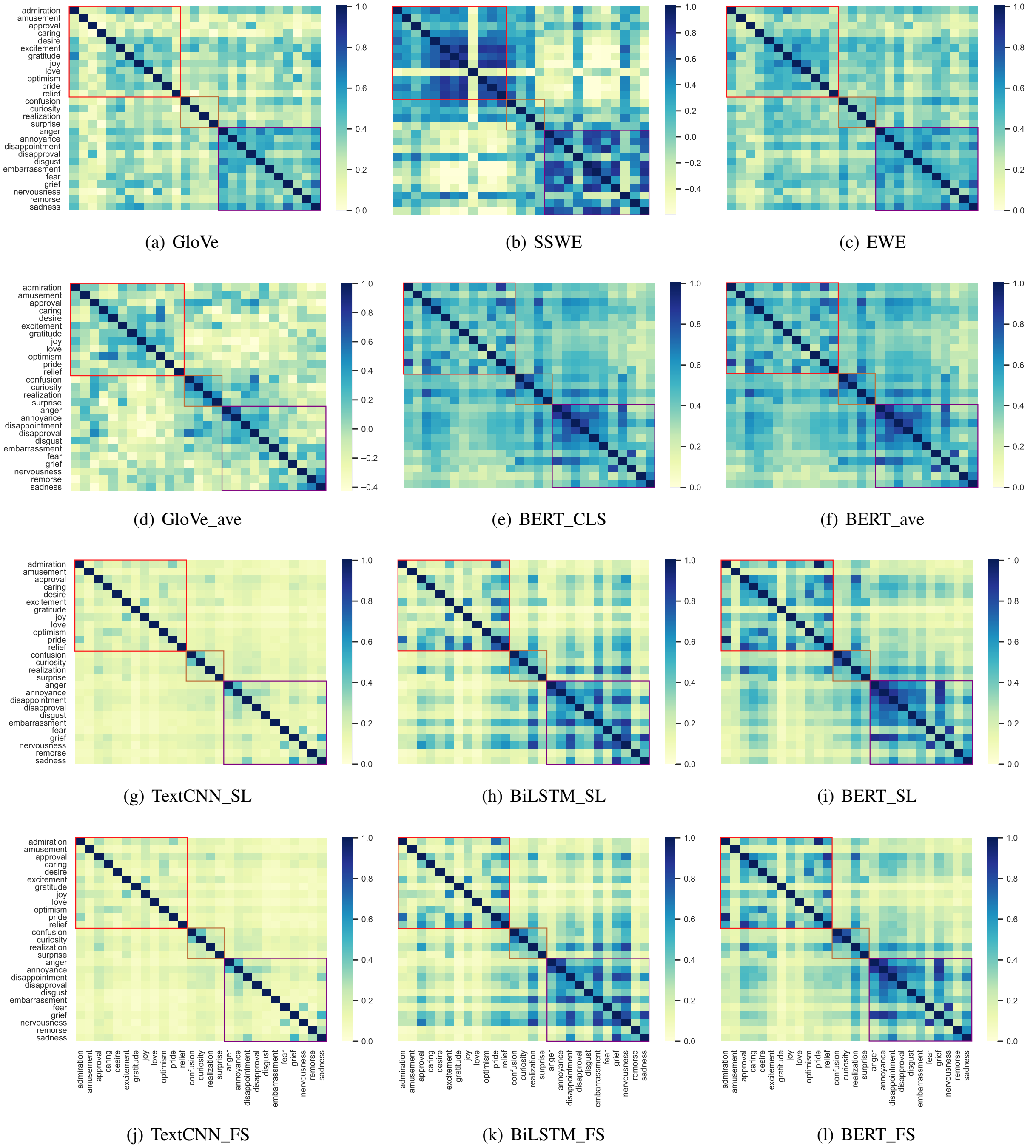


Fig. 5. The heatmaps show the cosine similarity between emotion vectors. Red, purple, and orange boxes correspond to sub-heatmaps of positive, negative and ambiguous emotion categories, respectively. (a)-(f): (emotion-enriched) semantic space. (g)-(l): emotion space. (SL: soft labels; FS: forced symmetry.)

ambiguous emotions[1].

The experiments were conducted 10 times. In each experiment, a distributed representation matrix for emotion categories was learned, and the average representation matrices were computed. Singular value decomposition (SVD) [62] was then applied to reduce the dimensionality of the average matrix from 27 to 2. SVD was chosen for its orthogonal transformation property, which preserves the spatial structure of the original data, ensuring that the reduced space retains the distribution and separability of the original. The resulting two-dimensional representation of emotion categories is shown in Figure 4.

Three color-shape pairs, red-circle, gray-square, and black-triangle, represent positive, negative, and ambiguous emotions, respectively. Figures 4(a)-(c) illustrate the word embeddings in semantic space (GloVe, SSWE, and EWE), while Figures 4(d)-(f) depict the averaged sentence embeddings (GloVe_ave, BERT_CLS, and BERT_ave). Figures 4(g)-(l) present the results of the DREC-SL and DREC-FS algorithms using three neural network models: TextCNN, BiLSTM, and BERT.

In contrast, Figures 4(e)-(l) display clear boundaries among

1. https://github.com/google-research/google-research/tree/master/goemotions/data/sentiment_mapping.json

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on March 19,2025 at 08:46:11 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2025.3545447

TABLE 4
Cosine similarities between emotion pairs. We also calculate the average similarity and standard deviations of similar and dissimilar emotion pairs. For similar emotion pairs, the highest average value and the smallest standard deviation are shown in bold. For dissimilar emotion pairs, the smallest positive average value and the smallest standard deviation are shown in bold.

		(Emotion-enriched) Semantic Space						**Emotion Space**					
		GloVe	**SSWE**	**EWE**	**GloVe ave**	**BERT CLS**	**BERT ave**	**TC SL**	**BL SL**	**BERT SL**	**TC FS**	**BL FS**	**BERT FS**
Similar	(anger,annoyance)	0.497	0.625	0.586	0.616	0.870	0.871	0.470	0.786	**0.883**	0.483	0.797	0.871
	(grief,sadness)	0.734	0.775	0.790	-0.005	0.602	0.607	0.560	**0.869**	0.656	0.503	0.849	0.640
	(joy,relief)	0.266	**0.824**	0.314	-0.164	0.711	0.715	0.266	0.806	0.754	0.232	0.819	0.753
	(fear,nervousness)	0.393	-0.293	0.442	0.317	0.817	0.817	0.414	0.725	0.855	0.376	0.671	**0.866**
	avg	0.473	0.483	0.533	0.191	0.750	0.753	0.427	**0.797**	0.787	0.398	0.784	0.783
	std	0.172	0.454	0.177	0.300	0.103	0.101	0.107	**0.051**	0.090	0.108	0.068	0.095
Dissimilar	(disgust,joy)	0.346	-0.779	0.304	-0.270	0.349	0.355	0.094	0.084	0.169	**0.077**	0.086	0.146
	(anger,admiration)	0.460	-0.100	0.428	-0.185	0.377	0.381	0.092	**0.054**	0.121	0.101	0.064	0.154
	(love,remorse)	0.191	-0.425	0.253	0.067	0.180	0.183	0.059	0.045	0.083	0.050	**0.041**	0.070
	(fear,gratitude)	0.188	-0.585	0.382	-0.376	0.275	0.277	0.040	0.043	0.066	0.043	**0.036**	0.061
	avg	0.296	-0.472	0.342	-0.191	0.295	0.299	0.071	**0.057**	0.110	0.068	**0.057**	0.108
	std	0.114	0.249	0.068	0.164	0.076	0.077	0.023	**0.016**	0.040	0.023	0.020	0.042

positive, negative, and ambiguous emotion categories. Notably, ambiguous categories are positioned between positive and negative categories, confirming their intermediate and ambiguous nature.

In Figure 4 (a), (b), and (d), there are no clear boundaries among positive, negative and ambiguous emotion categories. In Figure 4 (c), positive and negative emotions are separated with each other while ambiguous emotions seem to locate randomly in the space. This reflects that the similarity between emotion terms in semantic space cannot truly reflect relations between emotion categories. For example, *joy* and *sadness* are different emotion categories, but due to their similar context, these two emotions may have similar positions in semantic space.

In contrast, Figures 4(e)-(l) display clear boundaries among positive, negative, and ambiguous emotion categories. Notably, ambiguous categories are positioned between positive and negative categories, confirming their intermediate and ambiguous nature.

This experiment demonstrates that our emotion representation algorithm more effectively captures the relationships among positive, negative, and ambiguous emotion categories compared to traditional word vectors in semantic space.

5.4 Similarity

The similarities among emotion categories in emotion space will be calculated to quantitatively reveal emotion relations in this part. Cosine similarity is employed as the measurement of similarity in this paper.

5.4.1 Overall Similarities

The heatmap in Figure 5 illustrates the overall similarities among emotion categories. Positive, negative, and ambiguous emotions are highlighted with red, orange, and purple boxes, respectively. Among the models, SSWE (Figure 5(b)) exhibits the most distinct color depth differences between regions inside and outside the boxes. This is because SSWE incorporates positive and negative information of emotion terms as auxiliary data during training. In contrast, GloVe and EWE show no significant differences in color depth between these regions. For BERT_CLS and BERT_ave (Figures 5(e)-(f)), the overall darker color indicates a high cosine similarity coefficient across emotion categories.

In emotion space as shown in Figure 5 (g)-(l), the color depth inside the boxes is higher than that outside the boxes, which indicates the similar emotions tend to have a higher overall similarity. As for algorithm DREC-SL (Figure 5 (g)-(i)) and DREC-FS (Figure 5 (j)-(l)), the similarity heatmaps of the same neural model are very similar across algorithms.

5.4.2 Case Study

It should be noted that the human emotional state is actually a high-dimension space. Positive-negative is only one (but most important) dimension in emotion space. As a result, two emotion categories may have a low cosine similarity, even if they are both annotated as positive (or negative).

To assess our emotion representation's ability to capture relations among similar and dissimilar emotion pairs, we analyzed four pairs of each type. Cosine similarities and standard deviations were calculated over 10 runs with different initial parameters. For comparison, we also evaluated word vectors and three averaged sentence embeddings in semantic space. Key findings from Table 4 are as follows:

1. In emotion-enriched word embeddings (GloVe, SSWE, EWE) and GloVe_ave, dissimilar emotion pairs may exhibit higher cosine similarities than similar pairs. In contrast, BERT_CLS, BERT_ave, and our emotion space consistently yield higher cosine similarities for similar pairs. This difference stems from the training process: BERT_CLS and BERT_ave leverage embeddings from a fine-tuned BERT model, enabling them to better capture emotional distances.

2. Cosine similarities in emotion space are more distinct between similar and dissimilar pairs compared to semantic space. For instance, word embeddings may show negative similarity for similar pairs (e.g., *fear* and *nervousness* in SSWE) or medium similarity for dissimilar pairs (e.g., *anger* and *admiration* in GloVe). Averaged sentence embeddings (BERT_CLS and BERT_ave) can differentiate these pairs, but emotion space consistently assigns high similarity to similar pairs and low similarity to dissimilar pairs.

3. In emotion space, cosine similarities between dissimilar emotion pairs are slightly above zero, suggesting that positive and negative affect represent independent rather than opposing dimensions. This aligns with findings in psychological research [63].

The *similarity* experiment in this part shows that our representation algorithm in emotion space can better distinguish similar and dissimilar emotions, which cannot be achieved by traditional (emotion-enriched) word vectors.

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on March 19,2025 at 08:46:11 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2025.3545447

TABLE 5
The results of mapping Cowen taxonomy to Ekman taxonomy. The experiments are conducted for 10 times. And the average emotion representation matrix is employed to show the detailed mapping results. Human results are chosen as the gold answers and wrong results are marked in italic. All 10 emotion representation matrices are employed to show average results and standard deviations. The highest average value and the smallest standard deviation are shown in bold.

Emotions	**Human**	**(Emotion-enriched) Semantic Space**						**Emotion Space**						**Modern LLMs**		
		GloVe	**SSWE**	**EWE**	**GloVe ave**	**BERT CLS**	**BERT ave**	**TC_SL**	**BL_SL**	**BERT_SL**	**TC_FS**	**BL_FS**	**BERT_FS**	**GPT-4o**	**Deepseek**	**Qwen**
admiration	joy	*disgust*	joy	*disgust*	*disgust*	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy
amusement	joy	joy	joy	*disgust*	*anger*	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy
annoyance	anger	*disgust*	*disgust*	anger	anger	anger	anger	anger	anger	anger	anger	anger	anger	anger	anger	anger
approval	joy	*surprise*	*anger*	*surprise*	*surprise*	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy
caring	joy	*sadness*	joy	joy	joy	*fear*	*fear*	*sadness*	joy	*sadness*	*sadness*	joy	joy	joy	joy	*love*
confusion	surprise	*fear*	*joy*	*fear*	surprise	surprise	surprise	surprise	surprise	surprise	surprise	surprise	surprise	surprise	*fear*	surprise
curiosity	surprise	*fear*	surprise	surprise	surprise	surprise	surprise	surprise	surprise	surprise	surprise	surprise	surprise	surprise	*joy*	surprise
desire	joy	*fear*	joy	*fear*	joy	*disgust*	*disgust*	joy	joy	joy	joy	joy	joy	joy	joy	joy
disappointment	sadness	sadness	*fear*	sadness	*disgust*	*disgust*	*disgust*	sadness	sadness	sadness	sadness	sadness	sadness	sadness	sadness	sadness
disapproval	anger	*disgust*	*joy*	*disgust*	anger	*disgust*	*disgust*	anger	*disgust*	anger	anger	anger	anger	*disgust*	anger	anger
embarrassment	sadness	*anger*	*disgust*	*disgust*	sadness	*disgust*	*disgust*	*disgust*	*disgust*	*disgust*	*disgust*	*disgust*	*disgust*	sadness	*fear*	sadness
excitement	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy
gratitude	joy	*sadness*	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy
grief	sadness	sadness	*disgust*	sadness	*anger*	*anger*	*anger*	sadness	sadness	*anger*	sadness	sadness	*anger*	sadness	sadness	sadness
love	joy	joy	*fear*	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy
nervousness	fear	*anger*	*joy*	*sadness*	*sadness*	fear	fear	fear	fear	fear	fear	fear	fear	fear	fear	fear
optimism	joy	*fear*	joy	*sadness*	joy	*disgust*	*disgust*	joy	joy	joy	joy	joy	joy	joy	joy	joy
pride	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy
realization	surprise	*joy*	*joy*	*sadness*	surprise	surprise	surprise	surprise	*fear*	surprise	surprise	*fear*	surprise	surprise	surprise	surprise
relief	joy	*anger*	joy	*anger*	*fear*	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy	joy
remorse	sadness	sadness	*disgust*	sadness	sadness	sadness	sadness	sadness	sadness	sadness	sadness	sadness	sadness	sadness	sadness	sadness
Score	—	7	10	10	14	14	14	19	18	18	19	19	19	**20**	18	**20**
Avg(10 times)	—	—	—	—	—	14.700	14.700	18.200	17.500	17.900	18.400	18.100	**18.500**	—	—	—
Std(10 times)	—	—	—	—	—	2.002	2.002	1.166	0.671	0.831	0.663	1.044	**0.500**	—	—	—

5.5 Mapping

The mapping relations between Cowen's emotion taxonomy [9] and Ekman's emotion taxonomy [8] are detected in this part. As mentioned before, there are 27 emotion categories in Cowen's taxonomy, which contains all 6 basic emotion categories in Ekman's taxonomy. In this experiment, we regard remaining 21 emotion categories as source emotion, and map them into Ekman's taxonomy. In other words, we aim to find the most similar one from Ekman's six basic emotion categories for each source emotion. The formula is listed as follows:

$$e = \arg\max_{e_t} \mathrm{sim}(e_s, e_t), \tag{12}$$

where e_t is the emotion category in target emotions, e_s is the emotion category in source emotions and e is the mapping result of e_s. sim is the similarity function and the cosine similarity is selected here.

The emotion representation matrix was calculated 10 times with different random initial parameters, and the average matrix was used to present detailed mapping results. The score of the average matrix is shown in the third-to-last row of Table 5. Scores for individual matrices were also calculated, with their averages and standard deviations listed in the last two rows. Human mapping results, provided by the authors of GoEmotions, serve as the gold standard. [2]

GloVe correctly maps 7 out of 21 emotions, highlighting its limitation in capturing emotion relations without explicit emotional information. By incorporating sentiment (positive or negative) into word embeddings, SSWE [54] improves to correctly map 10 emotions, with 9 of these being *joy*, as positive emotions are easier to map due to their limited representation in Ekman's basic emotions. Similarly, EWE [55] maps 10 emotions, indicating that emotion-enriched embeddings still struggle to fully capture inter-emotion mapping relations.

2. https://github.com/google-research/google-research/tree/master/goemotions/data/ekman_mapping.json

In the emotion-enriched semantic space built on the GoEmotions dataset, all three methods (GloVe_ave, BERT_CLS, and BERT_ave) correctly map 14 out of 21 emotions. BERT_CLS and BERT_ave yield identical mapping results, with average scores of 14.7, outperforming GloVe_ave. This demonstrates that the trained BERT model more effectively captures emotional relations from the dataset.

In emotion space, all three neural models using DREC-SL and DREC-FS correctly map 18–19 emotion categories, with significantly higher scores than in semantic space. This clearly demonstrates the effectiveness of our algorithms in capturing emotion relations, which word embeddings in semantic space fail to achieve.

For the DREC-SL algorithm, BERT achieves a higher average score than TextCNN and BiLSTM, indicating its superior ability to extract emotion relations from the dataset. Additionally, BERT's smaller standard deviation demonstrates greater stability. For TextCNN and BiLSTM, the forced symmetry operation improves average scores from 18.2 and 17.5 to 18.4 and 18.1, respectively. However, the improvement for BERT is not obvious.

As shown in Table 5, five categories (*caring*, *disapproval*, *embarrassment*, *grief*, *realization*) differ from human mappings. *Caring* is mapped to *sadness* by three models but to *joy* by humans and other models. While inherently positive, *caring* often accompanies negative events [64], reflecting the distinct dimensions of *joy* and *sadness* [63]. For *embarrassment*, models map it to *disgust*, while humans map it to *sadness*, consistent with studies showing its ties to both emotions [65].

All models constructed in emotion space produce mappings largely consistent with human results, demonstrating the effectiveness of our algorithms in capturing emotion relations. However, the complexity of emotion relations means certain emotions (e.g., *embarrassment* to *disgust* and *sadness*, *caring* to *sadness* and *joy*) may align with multiple categories, suggesting no single correct mapping exists for some emotions.

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on March 19,2025 at 08:46:11 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2025.3545447


```
### Task Definition
Single-choice Question. Please select one emotion most similar to the given
emotion from the alternative emotion pool.
### Output Format
| Given Emotion | Emotion Pool | Your Selection |
| admiration | anger, disgust, fear, joy, sadness, surprise | your selection here |
| amusement | anger, disgust, fear, joy, sadness, surprise | your selection here |
| annoyance | anger, disgust, fear, joy, sadness, surprise | your selection here |
| approval | anger, disgust, fear, joy, sadness, surprise | your selection here |
| caring | anger, disgust, fear, joy, sadness, surprise | your selection here |
| confusion | anger, disgust, fear, joy, sadness, surprise | your selection here |
| curiosity | anger, disgust, fear, joy, sadness, surprise | your selection here |
| desire | anger, disgust, fear, joy, sadness, surprise | your selection here |
| disappointment | anger, disgust, fear, joy, sadness, surprise | your selection here |
| disapproval | anger, disgust, fear, joy, sadness, surprise | your selection here |
| embarrassment | anger, disgust, fear, joy, sadness, surprise | your selection here |
| excitement | anger, disgust, fear, joy, sadness, surprise | your selection here |
| gratitude | anger, disgust, fear, joy, sadness, surprise | your selection here |
| grief | anger, disgust, fear, joy, sadness, surprise | your selection here |
| love | anger, disgust, fear, joy, sadness, surprise | your selection here |
| nervousness | anger, disgust, fear, joy, sadness, surprise | your selection here |
| optimism | anger, disgust, fear, joy, sadness, surprise | your selection here |
| pride | anger, disgust, fear, joy, sadness, surprise | your selection here |
| realization | anger, disgust, fear, joy, sadness, surprise | your selection here |
| relief | anger, disgust, fear, joy, sadness, surprise | your selection here |
| remorse | anger, disgust, fear, joy, sadness, surprise | your selection here |
```

Fig. 6. Prompt for Emotion Mapping Experiment with LLMs.

5.6 Comparison with Modern LLMs

To demonstrate the current understanding of emotion category relations by state-of-the-art large language models, we conducted this experiment with GPT-4o, Deepseek-V3, and Qwen2.5-72B. All models were prompted with the same instruction, as shown in Figure 6. The results are listed in Table 5.

Upon analyzing the results, we observe that while the overall performance of these LLMs is generally high, there are notable differences in their mappings compared to human judgments. For instance, Qwen2.5-72B incorrectly maps *caring* to *love*, which is not present in the candidate emotion pool. This suggests a potential hallucination in the model, indicating a lack of adherence to the given instructions. However, it also objectively highlights the similarity between *caring* and *love*. *Caring* is inherently a positive emotion, often associated with love, but it can also accompany negative events. This dual nature aligns with the essence of *love*, which can manifest in various contexts, both positive and negative.

On the other hand, Deepseek-V3 slightly underperforms compared to the other models, misclassifying three emotions. This highlights the variability in performance across different LLMs. The advantage of our proposed method lies in its lightweight nature, which does not rely on extensive pre-trained data. Despite this, it achieves comparable results, demonstrating its effectiveness in accurately mapping emotions without the need for large-scale pre-training datasets.

5.7 Impact of Frequency

This subsection examines the impact of emotion frequency on category representations. As expected, higher-frequency categories tend to produce more stable results due to larger sample sizes during training. To test this, we used the BERT_FS algorithm with ten different initial parameters, calculating the average angular distance between vector representations for each category and correlating these distances with sample sizes. The dataset includes categories with sample sizes ranging from 57 to 4,130.

The relations between the average angular distance and category frequency is shown in Figure 7, with a red line indicating the least squares fit to the data points. Overall, the negative correlation

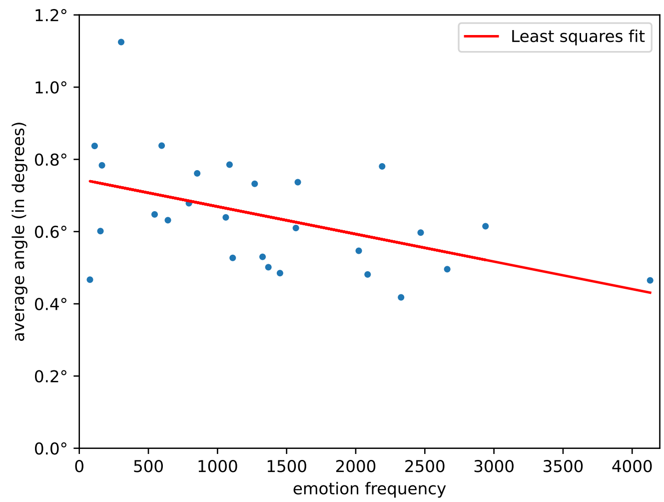


Fig. 7. The impact of frequency of available sentences per emotion category.

TABLE 6
Mapping Experiment Scores under Different Data Proportions

Data Proportion	10%	20%	40%	60%	80%
Mapping Score	14.6	14.2	17.2	17.6	18.3

coefficient between angle and frequency confirms that higher-frequency categories yield more stable results. Notably, even the category with the largest average angular distance had an angle of just 1.2 degrees, highlighting the robustness of the algorithm proposed in this study.

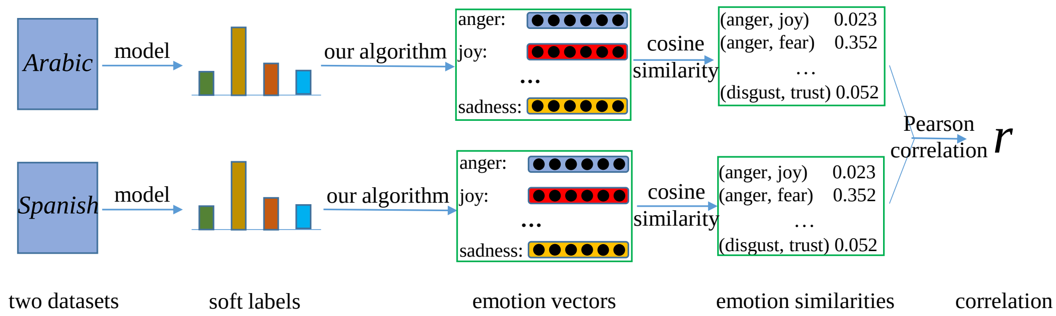
Table 6 presents the results of our mapping experiment under varying data proportions from 10% to 80%. Overall, the scores exhibit an upward trajectory as more data is introduced, suggesting that increased training samples generally enhance performance. Notably, even with just 10% of the data, the model achieves a mapping score of 14.6, indicating a commendable level of robustness in low-data scenarios. Beyond that, the performance rebounds significantly at 40% (17.2), and continues to improve, culminating in the highest score of 18.3 at 80%. These findings underscore both the method's capacity to leverage larger datasets for better outcomes and its resilience when the amount of available data is relatively small.

6 Application: Emotion Relations across Languages on Twitter—A Perspective from NLP

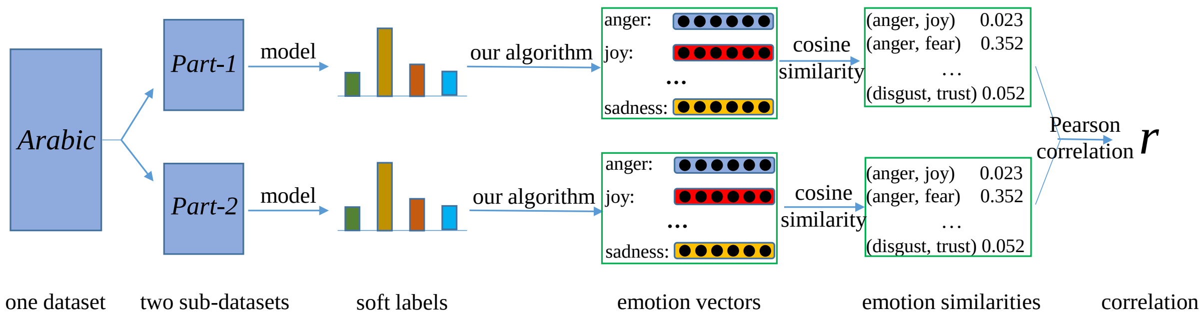
In this paper, we define the collection of emotion category relations in a language as the emotion structure of the language. In this section, we discuss emotion structures across languages on Twitter from an NLP perspective. In the previous one-hot representation, each emotion category is represented with an independent dimension. As a result, it is unable to analyze the emotion relations across languages with one-hot representation. However, based on the distributed representations of emotion categories proposed in this paper, we provide a new NLP perspective to study emotion relations across languages on Twitter.

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on March 19,2025 at 08:46:11 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2025.3545447



(a) Two different languages.



(b) The same language.

Fig. 8. Illustration of our approaches to detect emotion structures across languages. (a) Pearson correlation between emotion structures of two different languages. (b) Pearson correlation between emotion structures of the same language. (b) is employed to validate the reliability of our approaches.

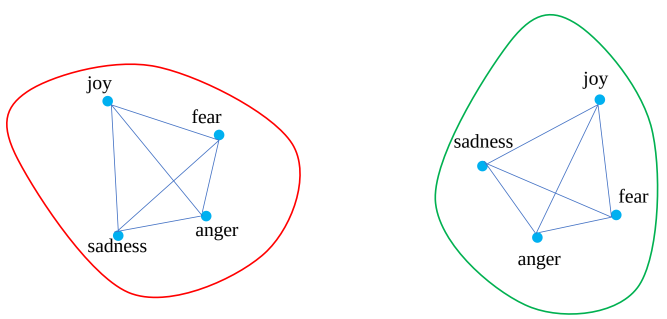


Fig. 9. Toy illustration of our approach. Red and green irregular shapes represent emotion structures conveyed through different languages, points represent the position of emotion categories, and lines between the points represent the relations between different emotion categories.

6.1 Model and Dataset

As validated in Section 5, our algorithms can learn a good emotion representation in emotion space with all three typical neural network models (TextCNN, BiLSTM and BERT). Since we focus on emotion relations across languages in this section, we choose BERT as the model for subsequent analysis. The BERT-based model of Arabic [66], English [60] and Spanish [67] are used in this part. The word vectors from reference [68] are chosen as multilingual word embedding.

The dataset we use in this section is presented in SemEval-2018 Task 1 [69] by Mohammad et al.. This dataset is the most comprehensive dataset for studying cross linguistic sentiment analysis, containing eleven emotion categories from three languages. The dataset is created from Twitter for the purpose of improving our understanding on how people convey emotions through language and contains three language-specific sub-datasets (English, Arabic, and Spanish). All three sub-datasets are collected from the same social media platform (Twitter) in the same time period and manually labeled with same classification criteria under same emotion taxonomy (*anger*, *anticipation*, *disgust*, *fear*, *joy*, *love*, *optimism*, *pessimism*, *sadness*, *surprise*, and *trust*). Therefore, this dataset can be employed not only to train a emotion classification model, but also to study emotion relations across languages on Twitter.

6.2 Approach

Relations between emotion categories are very complex [70], [71], and emotion structures conveyed through different languages are not the same [72]. As a result, it is difficult to directly measure the similarity between emotion structures of different languages. As shown in Figure 9, in order to analyze the differences between emotion structures conveyed through different languages, we regard emotion categories as the anchor points in emotion space. For each language, we calculate the cosine similarities among these emotion categories. Then, all the cosine similarities are get together to represent the emotion structures of the languages. Thus, the relations between emotion structures conveyed through different languages can be further detected.

The detailed approach can be seen in Figure 8. For each language, we learn the distributed representations for emotion categories from the soft labels output by pre-trained BERT model with our algorithm. Then, the similarity between each emotion pair are calculated. After that, all similarities are gathered together as the similarity vector of the language. Finally, the Pearson correlation coefficient between different similarity vectors of different languages are calculated as the measurement of similarity

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on March 19,2025 at 08:46:11 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2025.3545447

TABLE 7
Pearson correlation between different emotion structures. The experiments are conducted 10 times. The average results and standard deviations are listed in the Table. The maximum similarity within the same language (diagonal sections) and the maximum similarity across different languages (off-diagonal sections) are denoted in bold.

	DREC-SL			DREC-FS		
	Arabic	**English**	**Spanish**	**Arabic**	**English**	**Spanish**
Arabic	**0.987** (0.006)	0.720 (0.064)	0.778 (0.029)	**0.980** (0.012)	0.737 (0.077)	0.781 (0.029)
English	0.720 (0.064)	0.984 (0.020)	**0.846** (0.040)	0.737 (0.077)	0.982 (0.020)	**0.830** (0.041)
Spanish	0.778 (0.029)	**0.846** (0.040)	0.967 (0.016)	0.781 (0.029)	**0.830** (0.041)	0.958 (0.023)

TABLE 8
Significant p-value of the hypothesis that emotion structures of same language is higher than that of different language. For example, 2.43e-7 in first row and second column means the significant p-value of the hypothesis $H_0 : r(ar, ar) > r(ar, en)$ on t-test. 1.09e-6 in third row and first column means the significant p-value of the hypothesis $H_0 : r(es, es) > r(ar, es)$ on t-test. ar, en, es refer to Arabic, English and Spanish, respectively.

T-Test	DREC-SL			DREC-FS		
	Arabic	**English**	**Spanish**	**Arabic**	**English**	**Spanish**
Arabic	—	2.43e-7	2.09e-7	—	2.83e-6	5.46e-6
English	1.88e-9	—	3.38e-10	1.73e-9	—	7.83e-10
Spanish	1.09e-6	8.09e-7	—	5.62e-7	1.34e-6	—

TABLE 9
Significant p-value of the differences between emotion structures of different languages. For example, 7.17e-5 in first row and first column means the significant p-value of the hypothesis $H_0 : r(en, es) > r(ar, en)$ on t-test. ar, en, es refer to Arabic, English and Spanish, respectively.

T-Test	DREC-SL	DREC-FS
$r(en, es) > r(ar, en)$	7.17e-5	3.42e-3
$r(en, es) > r(ar, es)$	3.78e-4	4.48e-3
$r(ar, es) > r(ar, en)$	1.41e-2	7.07e-2

between different emotion structures. To analyze the potential error caused by the initial parameters of the model, we conduct the Pearson correlation coefficient for 10 times with different random initial parameters and record the average result and corresponding standard deviation. Moreover, in order to analyze the reliability of our approach, we randomly split each language-specific sub-dataset into two parts. And the similarity of the emotion structures of the two parts are calculated to compare with the desired result. (Ideally, the similarity of emotion structures of the two parts is supposed to be 1.)

6.3 Analysis

Table 7 summarizes the similarities and standard deviations of emotion structures across languages, highlighting both commonalities and differences. Here are the key findings:

High Similarity Across Languages: The Pearson correlation coefficients between emotion structures of different languages range from 0.720 to 0.846. This indicates that there is a significant degree of similarity in how emotions are structured and conveyed across Arabic, English, and Spanish. This finding aligns with the notion that human emotional experiences share commonalities, regardless of linguistic differences. This conclusion is consistent with previous studies in linguistics [73].

Significant Differences Exist: Despite the overall similarity, the Pearson correlation coefficients between different languages are notably lower than those within the same language, which range from 0.958 to 0.987. This suggests that while there are broad similarities, each language also has unique nuances in its emotion structure. These differences are statistically significant, as confirmed by t-tests, indicating that linguistic and cultural factors introduce distinct variations in emotional expression [74], [75].

Influence of Linguistic Families and Cultural History: The similarity between the emotion structures of English and Spanish is higher than that between English and Arabic, and higher than that between Spanish and Arabic. This can be attributed to the fact that English and Spanish both belong to the Indo-European language family [76], while Arabic belongs to the Semito-Hamitic language family [77]. Additionally, historical cultural exchanges have influenced the emotional lexicon and expressions between Arabic and Spanish, leading to a closer alignment in their emotion structures compared to English [78], [79], [80], [81].

In summary, emotions share universal features across languages, yet distinct linguistic and cultural contexts shape meaningful differences. This dual perspective provides a more comprehensive understanding of the intricate relationship between emotions and language, highlighting both the universality and the diversity of emotional expression across different linguistic communities.

7 Conclusion

In this paper, we propose a framework to learn distributed representations of emotion categories in emotion space. Then, two algorithms are derived on the basis of the soft labels predicted by the trained neural network model. Our algorithms represent emotion categories in emotion space rather than semantic space. Our algorithms also overcome the shortcomings of the previous one-hot representation that each emotion category is orthogonal to the others.

Comprehensive experiments confirm the efficiency and effectiveness of the proposed framework. *Arrangement* shows that positive, negative, and ambiguous emotions are linearly separated in emotion space, unlike in semantic space. *Similarity* demonstrates high cosine similarity for similar emotions and low for dissimilar

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on March 19,2025 at 08:46:11 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2025.3545447

ones in our representation. *Mapping* reveals that our framework accurately maps two emotion taxonomies, achieving results (18-19 out of 21) comparable to human performance, which semantic space embeddings cannot achieve.

For the first time, we explore cross-linguistic emotion structures on Twitter from the perspective of NLP with our proposed distributed representations of emotion categories. Our findings confirm insights from linguistics and psychology. First, while emotion structures show high similarity across languages, they are not identical. Second, English and Spanish have the most similar structures, as both belong to the Indo-European family, unlike Arabic, which is part of the Semito-Hamitic family. Third, Arabic and Spanish structures are more similar than Arabic and English, reflecting closer cultural ties between Arabic and Spanish cultures.

References

[1] L. F. Barrett, B. Mesquita, K. N. Ochsner, and J. J. Gross, "The experience of emotion," *Annu. Rev. Psychol.*, vol. 58, pp. 373–403, 2007.

[2] R. A. Calvo and S. Mac Kim, "Emotions in text: dimensional and categorical models," *Computational Intelligence*, vol. 29, no. 3, pp. 527–543, 2013.

[3] J. A. Russell, "A circumplex model of affect." *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

[4] A. Mehrabian and J. A. Russell, *An approach to environmental psychology.* the MIT Press, 1974.

[5] J. R. Fontaine, K. R. Scherer, E. B. Roesch, and P. C. Ellsworth, "The world of emotions is not two-dimensional," *Psychological science*, vol. 18, no. 12, pp. 1050–1057, 2007.

[6] S. Mohammad, "Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018, pp. 174–184.

[7] G. K. Verma and U. S. Tiwary, "Affect representation and recognition in 3d continuous valence–arousal–dominance space," *Multimedia Tools and Applications*, vol. 76, no. 2, pp. 2159–2183, 2017.

[8] P. Ekman, "An argument for basic emotions," *Cognition & emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.

[9] A. S. Cowen and D. Keltner, "Self-report captures 27 distinct categories of emotion bridged by continuous gradients," *Proceedings of the National Academy of Sciences*, vol. 114, no. 38, pp. E7900–E7909, 2017.

[10] R. Plutchik, "Emotions: A general psychoevolutionary theory," *Approaches to emotion*, vol. 1984, pp. 197–219, 1984.

[11] D. Keltner and J. S. Lerner, "Emotion," *Handbook of Social Psychology*, 2010.

[12] C. Harmon-Jones, B. Bastian, and E. Harmon-Jones, "The discrete emotions questionnaire: A new tool for measuring state self-reported emotions," *PloS one*, vol. 11, no. 8, p. e0159915, 2016.

[13] C. O. Alm, D. Roth, and R. Sproat, "Emotions from text: machine learning for text-based emotion prediction," in *Proceedings of human language technology conference and conference on empirical methods in natural language processing*, 2005, pp. 579–586.

[14] S. Y. M. Lee, Y. Chen, and C.-R. Huang, "A text-driven rule-based system for emotion cause detection," in *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, 2010, pp. 45–53.

[15] T. T. Thet, J.-C. Na, and C. S. Khoo, "Aspect-based sentiment analysis of movie reviews on discussion boards," *Journal of information science*, vol. 36, no. 6, pp. 823–848, 2010.

[16] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, "Target-dependent twitter sentiment classification," in *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies*, 2011, pp. 151–160.

[17] D. Zhou, X. Zhang, Y. Zhou, Q. Zhao, and X. Geng, "Emotion distribution learning from texts," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2016, pp. 638–647.

[18] S. Mohammad and F. Bravo-Marquez, "Emotion intensities in tweets," in *Proceedings of the 6th Joint Conference on Lexical and Computational Semantics (* SEM 2017)*, 2017, pp. 65–77.

[19] I. N. Alejo, T. Badia, and J. Barnes, "Cross-lingual emotion intensity prediction," *arXiv preprint arXiv:2004.04103*, 2020.

[20] R. Xia and Z. Ding, "Emotion-cause pair extraction: A new task to emotion analysis in texts," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 1003–1012.

[21] M. Abdalla and G. Hirst, "Cross-lingual sentiment analysis without (good) translation," in *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 2017, pp. 506–515.

[22] P. Dufter, M. Zhao, M. Schmitt, A. Fraser, and H. Schütze, "Embedding learning through multilingual concept induction," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018, pp. 1520–1530.

[23] M. Zhao and H. Schütze, "A multilingual bpe embedding space for universal sentiment lexicon induction," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 3506–3517.

[24] P. E. Ekman and R. J. Davidson, *The nature of emotion: Fundamental questions.* Oxford University Press, 1994.

[25] R. W. Levenson, "Basic emotion questions," *Emotion review*, vol. 3, no. 4, pp. 379–386, 2011.

[26] D. Keltner, D. Sauter, J. Tracy, and A. Cowen, "Emotional expression: Advances in basic emotion theory," *Journal of nonverbal behavior*, pp. 1–28, 2019.

[27] A. Wierzbicka, *Emotions across languages and cultures: Diversity and universals*. Cambridge University Press, 1999.

[28] R. B. Hupka, A. P. Lenton, and K. A. Hutchison, "Universal development of emotion categories in natural language." *Journal of personality and social psychology*, vol. 77, no. 2, p. 247, 1999.

[29] J. S. Boster, "Emotion categories across languages," in *Handbook of categorization in cognitive science*. Elsevier, 2005, pp. 313–352.

[30] P. Ekman and D. Keltner, "Universal facial expressions of emotion," *Segerstrale U, P. Molnar P, eds. Nonverbal communication: Where nature meets culture*, pp. 27–46, 1997.

[31] R. E. Jack, C. Blais, C. Scheepers, P. G. Schyns, and R. Caldara, "Cultural confusions show that facial expressions are not universal," *Current biology*, vol. 19, no. 18, pp. 1543–1548, 2009.

[32] X. Wang and C. Zong, "Distributed representations of emotion categories in emotion space," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Online: Association for Computational Linguistics, Aug. 2021, pp. 2364–2375. [Online]. Available: https://aclanthology.org/2021.acl-long.184

[33] B. Weiner and S. Graham, "An attributional approach to emotional development," *Emotions, cognition, and behavior*, pp. 167–191, 1984.

[34] R. Plutchik, "A general psychoevolutionary theory of emotion," in *Theories of emotion*. Elsevier, 1980, pp. 3–33.

[35] W. G. Parrott, *Emotions in social psychology: Essential readings*. psychology press, 2001.

[36] C. Strapparava and R. Mihalcea, "Semeval-2007 task 14: Affective text," in *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, 2007, pp. 70–74.

[37] W. Wang, L. Chen, K. Thirunarayan, and A. P. Sheth, "Harnessing twitter" big data" for automatic emotion identification," in *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*. IEEE, 2012, pp. 587–592.

[38] M. Abdul-Mageed and L. Ungar, "Emonet: Fine-grained emotion detection with gated recurrent neural networks," in *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: Long papers)*, 2017, pp. 718–728.

[39] Y. Li, H. Su, X. Shen, W. Li, Z. Cao, and S. Niu, "Dailydialog: A manually labelled multi-turn dialogue dataset," in *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 2017, pp. 986–995.

[40] E. Öhman, K. Kajava, J. Tiedemann, and T. Honkela, "Creating a dataset for multilingual fine-grained emotion-detection using gamification-based annotation," in *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 2018, pp. 24–30.

[41] K. R. Scherer and H. G. Wallbott, "Evidence for universality and cultural variation of differential emotion response patterning." *Journal of personality and social psychology*, vol. 66, no. 2, p. 310, 1994.

[42] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, and S. Ravi, "GoEmotions: A dataset of fine-grained emotions," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics, Jul. 2020, pp. 4040–4054.

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on March 19,2025 at 08:46:11 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2025.3545447

[43] C. E. Izard and S. Buechler, "Aspects of consciousness and personality in terms of differential emotions theory," in *Theories of emotion*. Elsevier, 1980, pp. 165–187.

[44] P. Ekman, W. V. Friesen, M. O'sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn, P. E. Ricci-Bitti *et al.*, "Universals and cultural differences in the judgments of facial expressions of emotion." *Journal of personality and social psychology*, vol. 53, no. 4, p. 712, 1987.

[45] D. Matsumoto, "Culture and emotion," *The Handbook of Culture and Psychology*, p. 171, 2001.

[46] S. A. Kotz and S. Paulmann, "Emotion, language, and the brain," *Language and Linguistics Compass*, vol. 5, no. 3, pp. 108–125, 2011.

[47] L. Sundararajan, "Understanding emotion in chinese culture," *Cham, Switzerland: Springer*, 2015.

[48] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.

[49] M. Phuong and C. Lampert, "Towards understanding knowledge distillation," in *International Conference on Machine Learning*, 2019, pp. 5142–5151.

[50] H. Zhang, L. Xiao, W. Chen, Y. Wang, and Y. Jin, "Multi-task label embedding for text classification," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 4545–4553.

[51] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[52] Y. Gao, W. Wang, C. Herold, Z. Yang, and H. Ney, "Towards a better understanding of label smoothing in neural machine translation," in *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, 2020, pp. 212–223.

[53] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[54] D. Tang, F. Wei, B. Qin, N. Yang, T. Liu, and M. Zhou, "Sentiment embeddings with applications to sentiment analysis," *IEEE transactions on knowledge and data Engineering*, vol. 28, no. 2, pp. 496–509, 2015.

[55] A. Agrawal, A. An, and M. Papagelis, "Learning emotion-enriched word representations," in *Proceedings of the 27th International Conference on Computational Linguistics*, 2018, pp. 950–961.

[56] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *International conference on machine learning*. PMLR, 2017, pp. 1321–1330.

[57] A. Berger, S. A. Della Pietra, and V. J. Della Pietra, "A maximum entropy approach to natural language processing," *Computational linguistics*, vol. 22, no. 1, pp. 39–71, 1996.

[58] Y. Kim, "Convolutional neural networks for sentence classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1746–1751.

[59] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.

[60] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 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)*, 2019, pp. 4171–4186.

[61] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.

[62] M. E. Wall, A. Rechtsteiner, and L. M. Rocha, "Singular value decomposition and principal component analysis," in *A practical approach to microarray data analysis*. Springer, 2003, pp. 91–109.

[63] F. A. Huppert and J. E. Whittington, "Evidence for the independence of positive and negative well-being: Implications for quality of life assessment," *British journal of health psychology*, vol. 8, no. 1, pp. 107–122, 2003.

[64] K. R. Scherer, V. Shuman, J. Fontaine, and C. Soriano Salinas, "The grid meets the wheel: Assessing emotional feeling via self-report," *Components of emotional meaning: A sourcebook*, 2013.

[65] K. R. Scherer, "What are emotions? and how can they be measured?" *Social science information*, vol. 44, no. 4, pp. 695–729, 2005.

[66] W. Antoun, F. Baly, and H. Hajj, "Arabert: Transformer-based model for arabic language understanding," in *LREC 2020 Workshop Language Resources and Evaluation Conference 11–16 May 2020*, 2020, p. 9.

[67] J. Cañete, G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, and J. Pérez, "Spanish pre-trained bert model and evaluation data," in *PML4DC at ICLR 2020*, 2020.

[68] E. Grave, P. Bojanowski, P. Gupta, A. Joulin, and T. Mikolov, "Learning word vectors for 157 languages," in *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.

[69] S. Mohammad, F. Bravo-Marquez, M. Salameh, and S. Kiritchenko, "Semeval-2018 task 1: Affect in tweets," in *Proceedings of the 12th international workshop on semantic evaluation*. New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 1–17. [Online]. Available: https://aclanthology.org/S18-1001

[70] L. F. Barrett and T. D. Wager, "The structure of emotion: Evidence from neuroimaging studies," *Current Directions in Psychological Science*, vol. 15, no. 2, pp. 79–83, 2006.

[71] M. De Choudhury, S. Counts, and M. Gamon, "Not all moods are created equal! exploring human emotional states in social media," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 6, no. 1, 2012.

[72] C. Goddard, *Explicating emotions across languages and cultures: A semantic approach*. Psychology Press, 2002.

[73] S. E. Kitayama and H. R. E. Markus, *Emotion and culture: Empirical studies of mutual influence.* American Psychological Association, 1994.

[74] B. Mesquita and H. R. Markus, "Culture and emotion," in *Feelings and emotions: The Amsterdam symposium*. Cambridge University Press, 2004, p. 341.

[75] J. M. Wilce and J. M. Wilce, *Language and emotion*. Cambridge University Press, 2009, no. 25.

[76] R. Bouckaert, P. Lemey, M. Dunn, S. J. Greenhill, A. V. Alekseyenko, A. J. Drummond, R. D. Gray, M. A. Suchard, and Q. D. Atkinson, "Mapping the origins and expansion of the indo-european language family," *Science*, vol. 337, no. 6097, pp. 957–960, 2012.

[77] R. Hetzron, "Afroasiatic languages," in *The world's major languages*. Routledge, 2017, pp. 561–566.

[78] R. Penny and R. J. Penny, *A history of the Spanish language*. Cambridge University Press, 2002.

[79] L. Quintana, J. P. Mora *et al.*, "Enseñanza del acervo léxico árabe de la lengua española," *Biblioteca virtual redELE*, 2004.

[80] D. Rorabaugh, "Arabic influence on the spanish language," *Research paper, Seattle Pacific University. https://mast. queensu. ca/˜ rorabaugh/docs/ArabicInfluence. pdf*, 2010.

[81] L. Sayahi, "Spanish in contact with arabic," *The Handbook of Hispanic Sociolinguistics*, pp. 473–489, 2011.

Xiangyu Wang received the ph.D. degree with the State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China, under the supervision of Prof. Chengqing Zong. He is currently with E Fund Management Co., Ltd, where his research interests include sentiment analysis and NLP applications.

Chengqing Zong received his Ph.D. from the Institute of Computing Technology, Chinese Academy of Sciences (CAS), in March 1998. From May 1998 to April 2000, he worked as a postdoctoral research fellow at the National Laboratory of Pattern Recognition (NLPR), Institute of Automation, CAS. After completing his postdoctoral program, he has continued his work at NLPR to the present. He is currently a professor at the Institute of Automation, CAS.

His research interests include fundamental research in computational linguistics, natural language processing, and cognitive language computing. As a principal investigator, he has led more than 10 projects supported by the Natural Science Foundation of China, the National Key Research and Development Program of China, and others. He has authored or co-authored six books and more than 200 journal and conference papers. In 2025, he is serving as the President of the Association for Computational Linguistics. He served as Program Committee Co-Chair for top-tier international conferences such as ACL 2015 and COLING 2020, and as General Conference Chair for ACL 2021. He is IEEE Fellow, ACL Fellow, CAAI Fellow and CCF Fellow.

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on March 19,2025 at 08:46:11 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.