Emotion Wheel Attention-Based Emotion Distribution Learning

XUEQIANG ZENG, QIFAN CHEN, XUEFENG FU, AND JIALI ZUO

1School of Computer and Information Engineering, Jiangxi Normal University, Nanchang, Jiangxi 330022, China
2School of Information Engineering, Nanchang Institute of Technology, Nanchang, Jiangxi 330099, China

Corresponding author: Jiali Zuo (zjl@jxnu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61866017, Grant 61866018, Grant 61762063, and Grant 61966019; and in part by the Jiangxi Natural Science Foundation under Grant 20192BAB207027.

ABSTRACT Emotion distribution learning is an effective multi-emotion analysis model proposed in recent years. Its core idea is to record the expression degree of examples on each emotion through emotion distribution, which is suitable for handling emotion analysis tasks with emotional ambiguity. To solve the problem that the prior knowledge of emotion psychology is seldom considered in the existing emotion distribution learning methods, we propose an Emotion Wheel Attention based Emotion Distribution Learning (EWA-EDL) model. EWA-EDL generates a prior emotion distribution describing the relevance of emotional psychology for each basic emotion, and then directly integrates the prior knowledge based on the emotion wheel into the deep neural network through the attention mechanism. The deep network of EWA-EDL is trained using an end-to-end approach to learn both emotion distribution prediction and emotion classification tasks. The EWA-EDL architecture includes five main parts: input layer, convolutional layer, pooling layer, attention layer and multi-task loss layer. Extensive comparative experiments on 8 commonly used textual emotion datasets show that EWA-EDL outperforms the comparison emotion distribution learning methods on both emotion distribution prediction and emotion classification task.

INDEX TERMS Emotion distribution learning, emotion wheel, attention mechanism, emotion classification.

I. INTRODUCTION

The task goal of emotion analysis is to uncover the emotional tendencies of people embedded in data [1], [2], which has important applications in many emerging artificial intelligence scenarios such as personalized recommendation [3] and intelligent customer service systems [4]. Emotion analysis mainly consists of three subtasks, namely emotion information extraction, emotion information classification, and emotion information retrieval and summarization [2]. The research in this paper focuses on emotion information classification, with the goal of improving the generalization performance of emotion recognition models.

Traditional emotion recognition models, mostly based on Single-Label Learning (SLL) or Multi-Label Learning (MLL), associate one or more emotion labels for examples and cannot quantitatively analyze multiple emotions with different expression strengths [5]. To address this problem, Zhou et al first proposed Emotion Distribution Learning (EDL) in 2015 for the facial expression recognition task [6]. EDL draws on Label Distribution Learning (LDL) [5], [7] to argue that emotions expressed by media such as text and images are a mixture of multiple basic emotions, with various basic emotions expressed at different intensities on the same example. The expression intensity of each basic emotion of an example is between 0 and 1, and the sum of the expression degrees of all emotions is 1. The expression degrees of all basic emotions on an example together constitute an Emotion Distribution [6]. In text emotion recognition tasks, it is a common phenomenon that a sentence expresses multiple basic emotions of different intensities at the same time. As shown in Figure 1, The SemEval text emotion dataset [8] gives specific scores for each sentence on how well it expresses each of the six basic emotions. EDL represents multiple emotions simultaneously through emotion distribution, and can better handle emotion recognition tasks where emotional ambiguity exists.

Human emotions are a complex phenomenon, with a variety of basic emotions highly interrelated, showing positive or negative correlations. Positively correlated emotions are more
likely to occur together, while negatively correlated emotions rarely occur together [9]. For example, example sentence (a) in Figure 1 labels two basic emotions, sadness and fear, where sadness is the dominant emotion. By manually analyzing the text content, we found that the example sentence (a) should also imply the emotion of surprise (marked with a dashed box in Figure 1). Further analysis reveals that surprise and sadness are two emotions that are highly positively correlated in psychology. The case of example sentence (b) is similar to that of example sentence (a), and the unlabeled aversive emotion highly correlated with sadness in this sentence can be found by manual analysis. How to effectively consider the correlation between emotions in emotion analysis models is a common research idea for various emotion recognition models.

In recent years, EDL is a research hotspot in the field of machine learning, and many scholars at home and abroad have published several EDL-related works in top international conferences and journals [10]–[14]. From the existing research work, it is an important research direction to consider the correlation between emotion labels in the EDL model. For example, Jia et al. [10] proposed a method for learning facial emotion distribution based on local relevance of labels using local low-rank structure to capture local relevance of labels in 2019. Xu and Wang [11] in 2021 proposed an attention mechanism-based method for learning the emotion distribution to obtain the relationship between various regions of an image and the emotion distribution. The basic idea of the above work is to add consideration to the emotional correlation mined from the training data in the EDL model, and has achieved certain results. However, there is still relatively little work on introducing emotional correlation based on prior knowledge of psychology into the EDL model.

Plutchik’s wheel of emotions is a classic psychological model of emotion proposed by psychologist Robert Plutchik in 1980 to describe the correlation between basic human emotions [15]. Plutchik considered human emotions as a mixture of eight basic emotions (anger, anticipation, joy, trust, fear, surprise, sadness, and disgust). The eight basic emotions together form Plutchik’s wheel of emotions, with two emotions in adjacent positions having a positive correlation and two emotions in opposite positions having a negative correlation. Overall, the interval angle on the emotion wheel represents the degree of psychological correlation between the corresponding emotions. He and Jin proposed a graph convolutional neural network-based EDL approach considering psychological prior knowledge on image emotion recognition task in 2019 with good results [12]. However, the model structure of this work is relatively simple and the training of the graph convolutional network containing prior knowledge is relatively independent. So far, no EDL work has adopted the attention mechanism to directly integrate the prior knowledge of psychology based on the emotion wheel into the deep learning network.

To effectively introduce psychological prior knowledge into the EDL model to improve the performance of emotion analysis, we propose the Emotion Wheel Attention based Emotion Distribution Learning (EWA-EDL) model in this paper. The EWA-EDL model is based on the classical textual convolutional neural network, and introduces prior knowledge of emotion relevance based on psychology into the neural network through the attention mechanism, and then uses a multi-task loss function to learn both the emotion distribution prediction and emotion classification tasks. The EWA-EDL model consists of five main components, namely, input layer, convolutional layer, pooling layer, attention layer, and multi-task loss layer. The input sentences are represented as a matrix of word embedding vectors in the input layer, and then new output features are generated using multiple filters of different widths in the convolution layer, followed by a standard maximum pooling operation in the pooling layer to obtain the maximum value among the features. In the attention layer, the EWA-EDL model constructs a prior emotion distribution
based on Plutchik’s wheel of emotions for each basic emotion, and fuses the prior emotion distribution of each emotion into a final emotion distribution output through the attention mechanism. In the multi-task loss layer, the EWA-EDL model combines KL (Kullback-Leibler) loss and cross-entropy loss to learn both emotion distribution prediction and emotion classification tasks simultaneously. In this paper, we conduct comparative experiments on English emotion distribution dataset (SemEval 2007 Task 14 [8]), four English single-label datasets (CBET [16], ISEAR [17], TEC [18] and Fairy Tales [19]) and three Chinese single-label datasets (NLP&CC 2013, NLP&CC 2014 [20] and WEC [21]). Experimental results show that the EWA-EDL model outperforms the comparison EDL model on both emotion distribution prediction and emotion classification tasks. The main contributions of the paper include the following two points:

(1) The attention mechanism is used to directly integrate the prior knowledge of psychology based on emotion wheels into the deep learning network, and proposed the EWA-EDL model.

(2) Comparative experiments on 8 commonly used text emotion data sets show that the EWA-EDL model proposed in this paper is fully compared with other EDL models. Experimental results show that the performance of EWA-EDL model on emotion distribution prediction and emotion classification tasks is better than the existing emotion distribution learning methods.

The remainder of this paper is organized as follows, Section 1 introduces emotion distribution learning and some related research work; Section 2 describes in detail emotion wheel attention based emotion distribution learning proposed in this paper; Section 3 describes the experimental setup, experimental results and analysis; and finally Section 4 concludes the full paper.

II. RELATED WORK

Traditional emotion analysis of texts focuses on classifying sentences in terms of emotion polarity, i.e., discerning positive, negative, or neutral polarity of emotion [2]. However, the emotion polarity recognition model is unable to capture fine-grained emotions and is only suitable for simple emotion analysis tasks. Unlike traditional emotion polarity analysis, the goal of fine-grained emotion analysis is to identify fine-grained emotions in a text [6], that is, specific emotions such as anger, anticipation, and joy. Classical fine-grained emotion recognition models are generally modeled based on single-label learning or multi-label learning, associating one or more emotion labels for examples [5]. Fine-grained emotion recognition models are capable of handling many emotion analysis tasks, but their modeling capabilities are still lacking to quantitatively represent the multiple emotions with varying degrees of expression that examples imply [13]. In practical applications, many sentences exist to express multiple emotions of varying degrees at the same time. For example, the commonly used SemEval text emotion dataset is annotated with the degree of expression of each of the six fine-grained emotions of the sentence [8]. As shown in Figure 1, sadness is the main emotion of sentence (a), with an expression level of 91.1%; as a secondary emotion, the expression level of fear is 8.82%. The degree of expression of the primary emotion (sadness) in sentence (b) is 82.85% and the degree of expression of the secondary emotion (anger) is 17.14%.

In order to quantitatively deal with the situation where an example expresses multiple emotions of different intensities at the same time, Zhou et al. [6] proposed Emotion Distribution Learning (EDL) in 2015, drawing on the research idea of Label Distribution Learning (LDL) [5]. Thereafter, Zhou et al. [13] proposed a text-oriented EDL method in 2016. A sentence may express one or more emotions, and each emotion has a different intensity of expression.

Emotion distribution learning can effectively handle the problem of expressing multiple emotions simultaneously in one example and is suitable for tasks where emotional ambiguity exists [6]. In recent years, many scholars have put forward much effective work in the field of EDL [11], [12], [14], [22]–[24]. Existing EDL works can be divided into three main categories according to whether or not they consider correlations between emotion labels.

The first class is EDL methods that do not consider the relevance of emotion labels, e.g., Zhang et al. [22] proposed an EDL model based on Multi-Task Convolutional Neural Network (MT-CNN) to optimize emotion distribution prediction and emotion classification tasks; Pandeya and Lee [23] presents an affective computing system that relies on music, video, and facial expression cues, making it useful for emotional analysis. And then Pandeya and Lee [24] construct a balanced music video emotion dataset including diversity of territory, language, culture and musical instruments.

The second class of methods learns label relevance from training data, which is a popular research direction in current EDL methods. Xu and Wang [11] proposed an attention mechanism-based method for learning emotion distribution, obtaining the relationship between various regions of an image and emotion distribution. Fei et al. [14] proposed a latent emotional memory network to learn the potential emotion distribution in the data, and used it effectively in the classification network.

The third class of EDL work considers the correlation between emotion labels using psychological emotional prior knowledge to improve the generalization performance of emotion recognition models. There is relatively little work on EDL considering prior knowledge of emotions, and He and Jin [12] proposed an EDL approach (EmotionGCN) based on an emotion wheel psychological model of emotion graph convolutional networks. EmotionGCN consists of two main components: a CNN module for extracting image features and a GCN-based weight generator that considers prior knowledge. However, the model structure of EmotionGCN is relatively simple. The GCN module, which learns the relevance of emotion labels, and the image-based CNN module are relatively independent, and only matrix multiplication
is used to merge the outputs of the two modules directly together. Researchers need to propose more effective EDL work that integrates psychological prior knowledge directly into deep learning networks. Compared with MT-CNN and EmotionGCN, the EWA-EDL model proposed in this paper is based on the emotion wheel psychology model to define the distance between each basic emotion and generate the corresponding prior emotion distribution. The distribution is fused with the convolutional neural network model, and then the multi-task loss function is used to simultaneously learn emotion distribution prediction and emotion classification tasks in an end-to-end manner.

III. EMOTION WHEEL ATTENTION-BASED EMOTION DISTRIBUTION LEARNING

A. PLUTCHIK’S WHEEL OF EMOTIONS

According to psychological research, human emotions are clearly interrelated [15]. Certain emotions often appear at the same time, that is, showing a high positive correlation; while other emotions are the opposite. The theory of the emotion wheel, proposed by Robert Plutchik in 1980, is a classic model for describing the interrelationship between emotions from a psychological perspective [15]. Plutchik’s wheel of emotions contains eight basic emotions: anger, disgust, sadness, surprise, fear, trust, joy, and expect. As shown in Figure 2, the eight basic emotions are divided into four groups of opposing emotions in the emotion wheel, with emotions in diagonal positions having negative correlations and those in adjacent positions having positive correlations.

![Figure 2. Plutchik’s wheel of emotions.](image)

In the field of EDL, there have been some researches using the emotion wheel as prior knowledge [12], [13]. For example, Zhou et al. [13] introduced an emotional constraint based on Plutchik’s wheel of emotions in the optimization objective of the EDL method based on the maximum entropy model and achieved good results. He and Jin [12] proposed an EDL method based on graph convolutional neural network and emotion wheel, which showed good performance. But overall, there has been relatively little work on EDL based on emotion wheels. At present, there is no EDL work that uses the attention mechanism to directly integrate the prior knowledge of psychology based on the emotion wheel into the deep learning network.

B. PLUTCHIK’S WHEEL OF EMOTIONS ATTENTION-BASED EMOTION DISTRIBUTION LEARNING

We use $d_i^j$ to denote the intensity of expression of the emotion $y$ in sentence $x$. The expressed intensity scores of various emotions on each sentence constitute an emotion distribution, which is guaranteed by vector normalization that $d_i^y \in [0, 1]$ and $\sum_y d_i^y = 1$. It is important to note that $d_i^y$ is not a probability, but the proportion of the emotion $y$ in the emotion distribution. If $d_i^y$ is considered as probability, it means that only one emotion label is correct for a sentence, while EDL considers an example to contain multiple emotions at the same time. The modeling goal of EDL is to learn a mapping from the sentence space $X = \mathbb{R}^M$ to the emotion distribution space $Y = \{y_1, y_2, \ldots, y_N\}$, with each label $y_i$ representing a basic emotion.

We introduce the psychological prior knowledge based on Plutchik’s Wheel of Emotions through the attention mechanism, and adopt a multi-task deep convolutional neural network to propose Emotion Wheel Attention based Emotion Distribution Learning (EWA-EDL) model. The architecture of the EWA-EDL model is shown in Figure 3.

Given a training dataset $S = \{(s_i, d_i)\}_{i=1}^N$ consisting of $N$ sentences, where $d_i = \{d_i^1, d_i^2, \ldots, d_i^M\}$ is the emotion distribution corresponding to sentence $s_i$. $d_i^j$ denotes the degree of expression of the $j$-th emotion on sentence $s_i$, $d_i^j \in [0, 1]$ and $\sum_j d_i^j = 1$. The modeling goal of EWA-EDL is to learn a mapping from the sentence $s_i$ to the emotion distribution $d_i$. EWA-EDL simultaneously optimizes two training tasks, namely emotion distribution prediction and emotion classification. The model architecture of EWA-EDL consists of five parts, namely, input layer, convolutional layer, pooling layer, attention layer, and multi-task loss layer.

1) INPUT LAYER

The input to the EWA-EDL model is a sentence $s = \langle w_1, w_2, \ldots, w_M \rangle$ consisting of $M$ words. We denote the $m$-th word $w_m$ as a $k$-dimensional word embedding vector $x_m \in \mathbb{R}^k$. Then, all the word vectors in the sentence are joined together to form the word vector matrix in Eq. 1:

$$x = [x_1, x_2, \ldots, x_M]$$  (1)
If the length of the sentence text does not reach $M$, we fill in the end of the word vector with 0 and represent each sentence as a word vector matrix of dimension $k \times M$.

2) CONVOLUTIONAL LAYER
The convolutional layer contains multiple filters $\omega \in \mathbb{R}^{h \times k}$, where $h$ is the window width, and each filter produces a new feature representation. Let $x_{p:p+h-1}$ denote the concatenation of words $x_p, x_{p+1}, \ldots, x_{p+h-1}$, then the feature $v_p$ is computed using filter $\omega$ on a set of words $x_{p:p+h-1}$, with the following equation in Eq. 2:

$$v_p = f(\omega \cdot x_{p:p+h-1} + b)$$

where $f(\cdot)$ is a nonlinear activation function, such as the Sigmoid or ReLu function, and $b$ is a bias term. The filter window covers all words $x_{1:h}, x_{2:h+1}, \ldots, x_{M-h+1:M}$ in the sentence, which can produce a set of feature maps $v$ in Eq. 3:

$$v = [v_1, v_2, \ldots, v_{M-h+1}]$$

3) POOLING LAYER
A series of standard max-over-time pooling operations applied to the feature map $v$ are used to obtain the maximum value in the feature as the significant feature in Eq. 4:

$$\hat{v} = \max(v)$$

where $\hat{v}$ denotes the feature representation corresponding to the input sample and the particular filter.

4) ATTENTION LAYER
The attention mechanism was first proposed in the field of computer vision, the main purpose is to allow the neural network to focus on specific parts of the image as needed, rather than the overall image [25]. Bahdana et al. [26] successfully applied the attention mechanism in the field of natural language processing in 2014. In view of the outstanding performance of the attention mechanism on natural language processing tasks, we use the attention mechanism to introduce prior knowledge of emotional psychology into the EDL model based on deep networks. We first generate a prior emotion distribution describing the correlation between emotions for each basic emotion, train the CNN to obtain the preliminary emotion distribution as weights, and use the attention mechanism to weight the final emotion distribution for prediction.

According to the psychological distance of various emotions in the emotion wheel model, we generate a prior emotion distribution $f_a$ for each emotion $\alpha$, $\alpha \in \{1, 2, \ldots, C\}$. We believe that in the prior emotion distribution $f_a$, the value of the emotion label $\alpha$ should be the largest and the degree of expression is the highest, and the values of other emotions decrease as the distance from the emotion label $\alpha$ in the emotion wheel increases. Overall, the prior emotion distribution $f_a$ should be a symmetrically decreasing distribution centered on the emotion label $\alpha$. The prior emotion distribution generated for each basic emotion is a fixed value. According to the conclusion of the LDL-based face age prediction work of Geng et al. [27], we assume that the prior emotion distribution obeys the Gaussian distribution. Given the emotion label $\alpha$, we generate a prior emotion distribution $f_a$ based on a Gaussian distribution, calculated as follows in Eq. 5 and Eq. 6:

$$f_{\alpha} = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{|\alpha - \alpha|^2}{2\sigma^2}\right)$$

$$Z = \frac{1}{\sigma \sqrt{2\pi}} \sum_{\alpha} \exp\left(-\frac{|\alpha - \alpha|^2}{2\sigma^2}\right)$$
The output vector $\hat{v}$ of the pooling layer is transformed through a fully connected layer and Softmax activation function to obtain the preliminary emotion distribution layer $g = [g_1, g_2, \ldots, g_C]$, where $g_j$ is the preliminary predicted degree of expression of the $j$-th emotion. Then, we weighted the attention to a prior emotion distribution of each basic emotion with the preliminary emotion distribution $g$. Finally, the weighted prior emotion distributions are superimposed to output an emotion distribution $\hat{d}$ that implies prior knowledge of the emotion psychology in Eq. 7.

$$\hat{d} = g_1 f_1 + g_2 f_2 + \cdots + g_C f_C$$  (7)

where $f_j$ is the prior emotion distribution of the $j$-th emotion $g_j$.

As shown in Figure 4, the calculation process of the attention layer of the EWA-EDL model is explained in detail. Take the preliminary predicted emotion distribution $g = [0, 0.7, 0, 0.3, 0, 0]$ obtained by CNN as an example. We weighted the prior emotion distribution for each basic emotion. In this example, only two emotions, joy and surprise, are not expressed to a degree of 0. The corresponding weighted prior emotion distributions are $[2.82 \times 10^{-2}, 6.42 \times 10^{-1}, 2.82 \times 10^{-2}, 5.68 \times 10^{-4}, 2.39 \times 10^{-6}, 5.68 \times 10^{-4}]$ and $[1.35 \times 10^{-4}, 1.35 \times 10^{-4}, 7.00 \times 10^{-2}, 1.52 \times 10^{-1}, 7.00 \times 10^{-2}, 6.72 \times 10^{-3}]$. Finally, we fuse the weighted emotion distributions to obtain the final predicted emotion distribution of the attention layer $[2.83 \times 10^{-2}, 6.42 \times 10^{-1}, 9.82 \times 10^{-2}, 1.53 \times 10^{-1}, 7.00 \times 10^{-2}, 7.28 \times 10^{-3}]$, which is also the final output of the EWA-EDL model.

5) MULTI-TASK LOSS LAYER

The EWA-EDL model combines cross-entropy loss function and KL loss function to train both emotion distribution prediction and emotion classification tasks using an end-to-end approach. Two simultaneously trained learning tasks can reinforce each other and learn to obtain a more robust neural network model. For the dataset with labeled emotion distribution, the emotion with the highest degree of expression in the emotion distribution $d_j$ is used as the true
emotion label of sentence $s_i$ for emotion classification. For the single-label dataset without labeled emotion distributions, we used label enhancement [28] to extend the true emotion labels into emotion distributions.

The objective loss function of EWA-EDL is a weighted combination of cross-entropy loss and KL loss by Eq. 8.

$$E = (1 - \lambda)E_{cls} + \lambda E_{edl}$$

where $E_{cls}$ represents the cross-entropy loss used for the emotion classification task, $E_{edl}$ is the KL loss for the emotion distribution prediction task, and $\lambda$ is the weight parameter. Based on previous work [22], $\lambda$ is set to 0.7.

The cross-entropy loss maximizes the probability of the target label, an objective function commonly used for classification tasks, defined as

$$E_{cls} = -\frac{1}{N} \sum_i \sum_j 1(y_i = j) \ln \frac{e^{y_i}_j}{\sum_l e^{y_i}_l}$$

where $1(\delta) = 1$ when $\delta$ is true and 0 otherwise, $y_i$ is the true emotion label of sentence $s_i$, and $\{a^{(i)}_j | j = 1, 2, \ldots, C\}$ denotes the output value of sentence $s_i$ at the last layer.

For emotion distribution prediction, the KL loss metric measures the difference between the predicted and true distributions, as defined in Eq. 10:

$$E_{edl} = -\frac{1}{N} \sum_i \sum_j d_i(j) \ln \frac{e^{y_i}_j}{\sum_l e^{y_i}_l}$$

where $d_i(j)$ is the total loss of different categories of labels for sentence $s_i$.

The source code of the EWA-EDL model is released at https://github.com/zeakyop/EWA-EDL.

### IV. EXPERIMENTS

To examine the performance of the EWA-EDL model proposed in this paper, three sets of experiments were conducted to analyze the effect of the parameter $\sigma$ of the prior emotion distribution on the performance of the EWA-EDL model, to compare the emotion prediction performance of multiple EDL methods on the English and Chinese datasets and to compare the emotion classification performance of three deep network-based EDL models on seven single-label datasets.

#### A. DATASETS

In this paper, eight textual datasets were used for the experiments, namely the SemEval emotion distribution dataset [8] (SemEval 2007 Task 14), four single-label English datasets (CBET [16], ISEAR [17], TEC [18] and Fairy Tales [19]) and three single-label Chinese datasets (NLP&CC 2013 [20], NLP&CC 2014 [20], and WEC [21]). We present the details of all Chinese and English experimental datasets in Table 1, including the number of sentences for each emotion, the number of all sentences, and the average number of words per sentence.

The basic information of the five English text datasets used in this paper is as follows. The SemEval dataset [8] is an emotion distribution dataset labeled with multiple emotion expression intensities and contains 1250 news headlines, each with six emotion labels and their corresponding expression intensities, i.e., anger, joy, fear, surprise, sadness, and disgust. The SemEval dataset mainly collects news headlines from mainstream English newspapers such as New York Times, CNN, BBC, and Google News to collect news headlines. The CBET dataset [16] consists of 76,860 tweet texts containing 9 emotions and 8540 tweets for each emotion. We selected a total of 51240 tweets from the CBET dataset for the six emotions included in the Plutchik’s wheel of emotions (anger, joy, fear, surprise, sadness, and disgust). The ISEAR dataset [17] contains 7666 sentences and 7 emotion labels. The sentences are about people describing their life situations and experiences when they experience certain emotions (anger, joy, fear, sadness, disgust, shame, and guilt). The five emotions (anger, joy, fear, sadness, disgust) and 5431 sentences from the ISEAR dataset were selected for the experiments in this paper. The TEC dataset [18] consists of 21051 emotional tweets, each labeled with one of the six emotions, i.e., anger, joy, fear, happiness, surprise, and sadness. The Fairy Tales dataset [19] from 185 fairy tales 1204 English sentences were extracted with 5 emotions (anger, joy, fear, surprise and sadness) and each sentence was labeled with 1 emotion label.

For the Chinese datasets, the NLP&CC 2013 and NLP&CC 2014 single-label Chinese datasets [20] contain 32,185 sentences (10,552 emotional sentences and 21,633 non-emotional sentences, respectively). The main source of these sentences is Chinese news from Chinese mainstream newspapers such as People's Daily, China Daily, and Xinhua News.

### TABLE 1. Chinese and English experimental datasets.

| Datasets      | Number of Sentences of each emotion | Total sentence num. | Avg. word num. per sentence |
|---------------|-------------------------------------|---------------------|-----------------------------|
|               | anger  | disgust | sadness | surprise | fear  | joy  |               |                    |                    |
| SemEval       | 91     | 42      | 192     | 435      | 262   | 228  | 1250        | 6.4                |
| CBET          | 8540   | 8540    | 8540    | 8540     | 8540  | 8540 | 51240       | 15.0               |
| ISEAR         | 1087   | 1081    | 1083    | -        | 1090  | 1090 | 5431        | 21.7               |
| TEC           | 1555   | 761     | 3830    | 3849     | 2816  | 8240 | 21051       | 15.3               |
| Fairy Tales   | 216    | 264     | 114     | 166      | 444   | 1204 | 24.0        |                    |
| NLP&CC 2013   | 1146   | 2075    | 1562    | 472      | 186   | 2140 | 7581        | 22.9               |
| NLP&CC 2014   | 1899   | 3130    | 2478    | 820      | 299   | 2805 | 11431       | 23.3               |
| WEC           | 4562   | 4876    | 14052   | 1011     | 661   | 9959 | 35121       | 39.7               |

For the Chinese datasets, the NLP&CC 2013 and NLP&CC 2014 single-label Chinese datasets [20] contain 32,185 sentences (10,552 emotional sentences and 21,633 non-emotional sentences, respectively). The main source of these sentences is Chinese news from Chinese mainstream newspapers such as People’s Daily, China Daily, and Xinhua News.
of each evaluation metric over ten cross-validations is used to evaluate the final performance of the EDL model.

For the emotion distribution prediction task, we used six commonly used EDL metrics for evaluating the quality of the predicted emotion distribution, including Euclidean, Squares, Squared $\chi^2$, KL Divergence, Cosine, and Intersection [13]. For the emotion classification task, we used four classification evaluation metrics, namely Precision, Recall, F1-score, and Accuracy. In addition, the parameters of the EWA-EDL model were set as shown in Table 2.

**Table 2. EWA-EDL model parameter settings.**

| Parameter          | Set value |
|--------------------|-----------|
| Word vector dimension | 300       |
| Sliding Window     | (3, 4, 5) |
| activation function | Relu      |
| Optimizer          | SGD       |
| Learning Rate      | 0.01      |
| Dropout            | 0.5       |
| Batch Size         | 50        |

The experiments in this paper are run on a Lenovo workstation with the main hardware configuration of Intel Core i9-10900X 3.70GHz 10-core CPU and 128G RAM. The operating system is Ubuntu 18.04, and the deep learning framework is Pytorch 1.5.0.

**C. EFFECT OF THE PRIOR EMOTION DISTRIBUTION PARAMETER $\sigma$ ON THE PERFORMANCE OF EWA-EDL MODEL**

The standard deviation parameter $\sigma$ of the prior emotion distribution is used to control the dispersion of the emotion distribution generated from the ground-truth emotion labels, and is an important parameter affecting the performance of the EWA-EDL model. When $\sigma$ is larger, the emotion distribution is more dispersed and the Gaussian distribution curve is flatter; when $\sigma$ is smaller, the emotion distribution curve is thinner and taller. To analyze the effect of the standard deviation parameter $\sigma$ on the performance of the EWA-EDL model, we take the parameter $\sigma$ from 0 to 1 at every interval of 0.1 and record the changes of the corresponding Accuracy and Cosine metrics. The experimental results of the effect of the prior emotion distribution parameter $\sigma$ on the performance of the EWA-EDL model on the SemEval dataset are shown in Figure 5.

As shown in Figure 5, the EWA-EDL model on the emotion prediction task of the SemEval dataset, the evaluation metrics Accuracy and Cosine both reached the highest at $\sigma = 0.8$. When $\sigma$ takes values between 0 and 0.7, Accuracy and Cosine basically show a gradually increasing trend, indicating that when expanding individual emotion labels into emotion distributions, appropriately increasing the dispersion of the emotion distribution is beneficial to improve the performance of
As shown in Table 3, the EWA-EDL model proposed in this paper showed better performance than other EDL models on the SemEval dataset in general. The EWA-EDL model outperforms the six models SA-CPNN, AA-KNN, SA-LDSVR, SA-IIS, SA-BFGS, and AA-BP on all 10 metrics of the emotion distribution prediction and emotion classification tasks. The average ranking of EWA-EDL on different indicators is the first, and MT-CNN and TextCNN are the second and third, indicating that the deep learning method is better than the traditional method on the EDL task. On the emotion distribution prediction task, the EWA-EDL model showed optimal results on all five metrics except for KL Divergence, which was slightly lower than the MT-CNN model. For example, on the Cosine metric, the EWA-EDL is 0.9% higher than the next best MT-CNN model. The EWA-EDL model also showed excellent performance on the emotion classification task. For example, on the Accuracy metric, the EWA-EDL model outperformed the next best MT-CNN model by 1.68%. In the other three classification metrics except for Precision, the EWA-EDL model achieves the best score. The above experimental results show that the EWA-EDL model has a better performance of emotion prediction than the compared EDL model. Plutchik’s wheel of emotions can better describe the correlation between basic emotions, and the introduction of psychological prior knowledge based on the emotion wheel through the attention mechanism has a significant effect on improving the performance of the EDL model.

**FIGURE 5.** Effect of the prior emotion distribution parameter \( \sigma \) on the performance of the EWA-EDL model on the SemEval dataset.

The EDL model. When \( \sigma = 0.8 \), Cosine and Accuracy reach the highest value, indicating that the dispersion degree of the generated emotion distribution is optimal at this time. Taking surprise emotion as an example, the prior emotion distribution at \( \sigma = 0.8 \) is \([4.51 \times 10^{-4}, 4.51 \times 10^{-4}, 2.33 \times 10^{-1}, 5.09 \times 10^{-1}, 2.33 \times 10^{-1}, 2.24 \times 10^{-2}]\). When the value of \( \sigma \) increased to between 0.8 and 1, the evaluation indexes Cosine and Accuracy showed a significant decrease, indicating that the score of true emotion in the emotion distribution was too low at this time and the dispersion of the generated emotion distribution was too large. The above experimental results indicate that while maintaining the dominance of true emotion, appropriately increasing the dispersion degree of the prior emotion distribution helps to improve the performance of the EDL model.

### D. PERFORMANCE COMPARISON OF MULTIPLE EDL MODELS FOR EMOTION DISTRIBUTION PREDICTION AND EMOTION CLASSIFICATION

To evaluate the performance of the EWA-EDL model proposed in this paper for emotion distribution prediction and emotion classification on traditional emotion distribution datasets, we compare EWA-EDL with commonly used EDL and LDL methods, including SA-CPNN, AA-KNN, SA-LDSVR, SA-IIS, SA-BFGS, AA-BP [5], TextCNN [31] and MT-CNN [22]. AA-KNN and AA-BP are extended versions of the classical KNN algorithm and BP (Back Propagation) neural network to solve LDL tasks [5]. SA-LDSVR, SA-IIS, SA-BFGS and SA-CPNN are algorithms specifically designed for LDL tasks [5]. TextCNN [30] is a convolutional neural network model for text emotion classification. MT-CNN [22] is an algorithm proposed by Zhang et al. proposed a multi-task convolutional neural network model for textual EDL. The specific experimental results of eight emotion distribution learning methods for emotion distribution prediction and emotion classification on the SemEval dataset are shown in Table 3 (the best results for each metric are bolded).

As shown in Table 3, the EWA-EDL model proposed in this paper showed better performance than other EDL models on the SemEval dataset in general. The EWA-EDL model outperforms the six models SA-CPNN, AA-KNN, SA-LDSVR, SA-IIS, SA-BFGS, and AA-BP on all 10 metrics of the emotion distribution prediction and emotion classification tasks. The average ranking of EWA-EDL on different indicators is the first, and MT-CNN and TextCNN are the second and third, indicating that the deep learning method is better than the traditional method on the EDL task. On the emotion distribution prediction task, the EWA-EDL model showed optimal results on all five metrics except for KL Divergence, which was slightly lower than the MT-CNN model. For example, on the Cosine metric, the EWA-EDL is 0.9% higher than the next best MT-CNN model. The EWA-EDL model also showed excellent performance on the emotion classification task. For example, on the Accuracy metric, the EWA-EDL model outperformed the next best MT-CNN model by 1.68%. In the other three classification metrics except for Precision, the EWA-EDL model achieves the best score. The above experimental results show that the EWA-EDL model has a better performance of emotion prediction than the compared EDL model. Plutchik’s wheel of emotions can better describe the correlation between basic emotions, and the introduction of psychological prior knowledge based on the emotion wheel through the attention mechanism has a significant effect on improving the performance of the EDL model.

**E. EMOTION CLASSIFICATION PERFORMANCE COMPARISON OF DEEP NETWORK-BASED EMOTION DISTRIBUTION LEARNING MODELS**

To evaluate the performance of the EWA-EDL model on traditional single-label emotion datasets, we conducted experiments comparing the performance of emotion classification on four English single-label datasets (CBET, ISEAR, TEC and Fairy Tales) and three Chinese single-label datasets (NLP&&CC 2013, NLP&&CC 2014 and WEC). On the single-label datasets, we used the LLE (Lexicon-based emotion distribution Label Enhancement) label enhancement method [28] to enhance the example emotion labels into emotion distributions. The LLE method introduces lexicon information in the text to generate emotion distributions in addition to the true emotion labels of the sentences, which has a better good performance [28]. We compare the EWA-EDL model with deep neural network-based TextCNN [31] and MT-CNN model [22] for experiments, and the specific experimental results are shown in Table 4 (the best results of each metric are marked in bold).

As shown by the results in Table 4, the EWA-EDL model proposed in this paper exhibits an overall better performance than the TextCNN and MT-CNN models on the seven single-label datasets. Taking the F1 score and accuracy metrics as examples, the EWA-EDL model achieves the best scores on all seven datasets. Specifically, on the CBET, ISEAR, TEC, Fairy Tales, NLP&&CC 2013, NLP&&CC 2014, and WEC datasets, the F1 score of the EWA-EDL
TABLE 3. Experimental results comparing the performance of eight emotion distribution learning methods for task 1 (emotion distribution prediction) and task 2 (emotion classification) on the SemEval dataset (↑ indicates that the larger the indicator is, ↓ indicates that the smaller the indicator is).

| Method          | AA-BP | SA-BFGS | SA-IIS | SA-LSDVR | AA-KNN | SA-CPNN | TextCNN | MT-CNN | EWA-EDL |
|-----------------|-------|---------|--------|----------|--------|---------|---------|--------|---------|
| Task 1          |       |         |        |          |        |         |         |        |         |
| EuC (↑)         | 0.4512 (5) | 0.5428 (8) | 0.4803 (6) | 0.4488 (4) | 0.5102 (7) | 0.6292 (9) | 0.4227 (3) | 0.3340 (2) | 0.3278 (1) |
| Sre (↑)         | 0.4317 (4) | 0.4736 (7) | 0.4718 (6) | 0.4359 (5) | 0.4811 (8) | 0.5718 (9) | 0.4306 (3) | 0.3195 (2) | 0.3169 (1) |
| Squ (↑)         | 0.4354 (4) | 0.5029 (7) | 0.4921 (6) | 0.4261 (3) | 0.5287 (8) | 0.7005 (9) | 0.4678 (5) | 0.3626 (2) | 0.3607 (1) |
| K-L. (↑)        | 0.7360 (6) | 1.2851 (8) | 0.6870 (4) | 0.5980 (3) | 0.8197 (7) | 1.7785 (9) | 0.7347 (5) | 0.4104 (2) | 0.4079 (1) |
| Cos (↑)         | 0.7061 (5) | 0.6392 (7) | 0.6954 (6) | 0.7368 (4) | 0.6241 (8) | 0.5322 (9) | 0.7833 (3) | 0.8245 (2) | 0.8335 (1) |
| Int (↑)         | 0.5682 (3) | 0.5263 (6) | 0.5281 (5) | 0.5640 (4) | 0.5184 (7) | 0.4281 (9) | 0.5158 (8) | 0.6894 (2) | 0.6880 (1) |
| Task 2          |       |         |        |          |        |         |         |        |         |
| Pre. (↑)        | 0.3316 (5) | 0.3284 (6) | 0.2662 (7) | 0.2367 (8) | 0.3678 (4) | 0.2306 (9) | 0.3961 (3) | 0.4870 (1) | 0.4806 (2) |
| Rec. (↑)        | 0.3238 (4) | 0.3173 (5) | 0.2427 (8) | 0.2591 (6) | 0.1845 (9) | 0.2564 (7) | 0.3557 (3) | 0.4555 (2) | 0.4426 (1) |
| F1 (↑)          | 0.3167 (5) | 0.3177 (4) | 0.2213 (6) | 0.2096 (8) | 0.1280 (9) | 0.2202 (7) | 0.3425 (3) | 0.4385 (2) | 0.4426 (1) |
| Acc. (↑)        | 0.4400 (4) | 0.4040 (5) | 0.3960 (6) | 0.3760 (7) | 0.3640 (8) | 0.2560 (9) | 0.4920 (3) | 0.5496 (2) | 0.5664 (1) |
| Ave. Rank (↑)   | 4.5 (4) | 6.3 (7) | 6.0 (6) | 5.2 (5) | 7.5 (8) | 8.6 (9) | 3.9 (3) | 1.9 (2) | 1.1 (1) |

TABLE 4. Experimental results comparing the performance of three deep network-based emotion distribution learning models for emotion classification on seven single-label datasets.

| Datasets       | Models    | Pre(%)  | Rec(%)  | F1(%)  | Acc(%)  |
|----------------|-----------|---------|---------|--------|---------|
| CBET           | TextCNN   | 62.06   | 56.08   | 57.99  | 56.08   |
|                | MT-CNN    | 63.83   | 57.73   | 58.23  | 57.73   |
|                | EWA-EDL   | 63.71   | 59.38   | 59.48  | 59.36   |
| ISERAR         | TextCNN   | 71.22   | 68.26   | 68.07  | 67.14   |
|                | MT-CNN    | 72.20   | 69.94   | 69.77  | 69.92   |
|                | EWA-EDL   | 72.33   | 71.25   | 70.89  | 71.17   |
| TEC            | TextCNN   | 54.83   | 43.47   | 46.12  | 54.63   |
|                | MT-CNN    | 57.68   | 44.30   | 48.62  | 58.75   |
|                | EWA-EDL   | 58.39   | 46.12   | 49.69  | 59.41   |
| Fairy Tales    | TextCNN   | 74.18   | 58.75   | 61.60  | 68.16   |
|                | MT-CNN    | 76.13   | 59.17   | 62.85  | 69.66   |
|                | EWA-EDL   | 76.92   | 62.60   | 66.47  | 71.66   |
| NLP & CC 2013  | TextCNN   | 55.80   | 43.08   | 45.88  | 59.45   |
|                | MT-CNN    | 59.87   | 46.41   | 48.16  | 61.27   |
|                | EWA-EDL   | 61.10   | 46.09   | 48.65  | 62.08   |
| NLP & CC 2014  | TextCNN   | 54.63   | 42.88   | 45.46  | 58.70   |
|                | MT-CNN    | 58.64   | 49.71   | 51.73  | 60.41   |
|                | EWA-EDL   | 61.04   | 50.51   | 52.04  | 61.10   |
| WFC            | TextCNN   | 54.28   | 40.62   | 43.36  | 58.73   |
|                | MT-CNN    | 59.06   | 41.62   | 44.80  | 60.74   |
|                | EWA-EDL   | 59.74   | 41.68   | 44.97  | 61.04   |

In this paper, we propose an Emotion Wheel Attention based Emotion Distribution Learning (EWA-EDL) model by using the attention mechanism to effectively utilize psychological prior knowledge. The EWA-EDL model computes correlations among basic emotions based on Plutchik’s wheel of emotions, and incorporates psychological prior knowledge into a multi-task convolutional neural network with the attention mechanism. Comparative experimental results on eight Chinese and English text emotion datasets show that the EWA-EDL model has better performance than existing EDL methods on emotion distribution prediction and emotion classification tasks.

In future work, we will consider using prior knowledge in a more effective way to improve the attention mechanism and improve the performance of the emotion distribution learning model.

REFERENCES

[1] L. Zhang, S. Wang, and B. Liu, “Deep learning for sentiment analysis: A survey,” Wiley Interdiscip. Rev.: Data Mining Knowl. Discovery, vol. 8, no. 4, Mar. 2018, Art. no. e01253, doi: 10.1002/widm.1253.
[2] Y.-Y. Zhao, B. Qin, and T. Liu, “Sentiment analysis,” J. Softw., vol. 21, no. 8, pp. 1834–1848, 2010.
[3] M. Saraswat, S. Chakraverty, and A. Kala, “Analyzing emotion based movie recommender system using fuzzy emotion features,” Int. J. Inf. Technol., vol. 12, no. 2, pp. 467–472, Jun. 2020.
[4] Y. Song, C. Wang, C. Chen, W. Zhou, and H. Chen, “Sentiment analysis for intelligent customer service chatbots,” J. Chin. Inf. Process., vol. 34, no. 2, pp. 80–95, Feb. 2020.

[5] X. Geng, “Label distribution learning,” IEEE Trans. Knowl. Data Eng., vol. 28, no. 7, pp. 1734–1748, Jul. 2016.

[6] Y. Zhou, H. Xue, and X. Geng, “Emotion distribution recognition from facial expressions,” in Proc. 23rd ACM Int. Conf. Multimedia, Brisbane, VIC, Australia, Oct. 2015, pp. 1247–1250.

[7] X. Geng, S. Luo, S. Chen, S. Wu, and Z. Wan, “The facial age estimation based on adaptive multivariate multiple regression,” J. Jiangxi Normal Univ. (Natural Sci. Ed.), vol. 43, no. 1, pp. 68–75, Jan. 2019.

[8] C. Strapparava and R. Mihalcea, “SemEval-2007 task 14: Affective text,” in Proc. 4th Int. Workshop Semantic Eval., Prague, Czech Republic, Jun. 2007, pp. 70–74.

[9] Z. Xu, S. Wang, and C. Wang, “Exploiting multi-emotion relations at feature and label levels for emotion tagging,” in Proc. 28th ACM Int. Conf. Multimedia, San Francisco, WA, USA, Oct. 2020, pp. 2955–2963.

[10] X. Jia, X. Zheng, W. Li, C. Zhang, and Z. Li, “Facial emotion distribution learning by exploiting low-rank label correlations locally,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Long Beach, CA, USA, Jun. 2019, pp. 9841–9850.

[11] Z. Xu and S. Wang, “Emotional attention detection and correlation exploration for image emotion distribution learning,” IEEE Trans. Affect. Comput., early access, Apr. 6, 2021, doi: 10.1109/TAFFC.2021.3071131.

[12] T. He and X. Jin, “Image emotion distribution learning with graph convolutional networks,” in Proc. Int. Conf. Multimedia Retr., Ottawa, ON, Canada, Jun. 2019, pp. 382–390.

[13] D. Zhou, X. Zhang, Y. Zhou, Q. Zhao, and X. Geng, “Emotion distribution learning from texts,” in Proc. Conf. Empirical Methods Natural Lang. Process., Austin, TX, USA, 2016, pp. 638–647.

[14] H. Fei, Y. Zhang, Y. Ren, and D. Ji, “Latent emotion memory for multi-label emotion classification,” in Proc. AAAI Conf. Artif. Intell. (AAAI), New York, NY, USA, Feb. 2020, pp. 7692–7699.

[15] R. Plutchik, “A general psychoevolutionary theory of emotion,” Theories Emotion, vol. 1, no. 3, pp. 3–33, 1980.

[16] A. Ghollipour Shahraki, “Emotion mining from text,” M.S. thesis, Dept. Comput. Sci., Alberta Univ., Edmonton, AB, Canada, 2015.

[17] K. R. Scherer and H. G. Wallbott, “Evidence for universality and cultural variation of differential emotion response patterning,” J. Personality Social Psychol., vol. 66, no. 2, pp. 310–328, 1994.

[18] S. Mohammad, “# Emotional tweets,” in Proc. 1st Joint Conf. Lexical Comput. Semantics, Montreal, QC, Canada, Jun. 2012, pp. 246–255.

[19] C. O. Alm and R. Sproat, “Emotional sequencing and development in fairy tales,” in Proc. 1st Int. Conf. Affect. Comput. Intell. Interact., Beijing, China, Oct. 2005, pp. 668–674.

[20] Y. Yao, S. Wang, R. Xu, B. Liu, L. Gui, Q. Lu, and X. Wang, “The construction of an emotion annotated corpus on microblog text,” J. Chin. Inf. Process., vol. 28, no. 5, pp. 83–91, Jun. 2014.

[21] M. Li, Y. Long, Q. Lu, and W. Li, “Emotion corpus construction based on selection from hashtags,” in Proc. 10th Int. Conf. Lang. Resour. Eval., Portoro, Slovenia, May 2016, pp. 1845–1849.

[22] Y. Zhang, J. Fu, D. She, Y. Zhang, S. Wang, and J. Yang, “Text emotion distribution learning via multi-task convolutional neural network,” in Proc. 27th Int. Joint Conf. Artif. Intell., Stockholm, Sweden, Jul. 2018, pp. 4595–4601.

[23] Y. R. Pandeya, B. Bhattarai, and J. Lee, “Deep-learning-based multimodal emotion classification for music videos,” Sensors, vol. 21, no. 14, pp. 4927, Jul. 2021.

[24] Y. R. Pandeya and J. Lee, “Deep learning-based late fusion of multimodal information for emotion classification of music video,” Multimedia Tools Appl., vol. 80, no. 2, pp. 2887–2905, Jan. 2021.

[25] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, “Recurrent models of visual attention,” in Proc. Conf. Neural Inf. Process. Syst., Montreal, QC, Canada, Dec. 2014, pp. 2204–2212.

[26] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” 2014, arXiv:1409.0473. [Online]. Available: http://arxiv.org/abs/1409.0473

[27] X. Geng, C. Yin, and Z.-H. Zhou, “Facial age estimation by learning from label distributions,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 10, pp. 2401–2412, Oct. 2013.

[28] N. Xu, Y.-P. Liu, and X. Geng, “Label enhancement for label distribution learning,” IEEE Trans. Knowl. Data Eng., vol. 33, no. 4, pp. 1632–1643, Apr. 2021.

[29] T. Mikolov, I. Sutskever, K. Chen, and G. S. Corrado, “Distributed representations of words and phrases and their compositionality,” in Proc. 26th Conf. Neural Inf. Process. Syst., Sydney, NSW, Australia, Dec. 2013, pp. 3111–3119.

[30] S. Li, Z. Zhao, R. Hu, W. Li, T. Liu, and X. Du, “Analogical reasoning on Chinese morphological and semantic relations,” in Proc. 56th Annu. Meeting Assoc. Comput. Linguistics, vol. 2, Melbourne, VIC, Australia, 2018, pp. 138–143.

[31] Y. Kim, “Convolutional neural networks for sentence classification,” in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), Doha, Qatar, Oct. 2014, pp. 1746–1751.