MDPI

Lexicon-Enhanced Multi-Task Convolutional Neural Network for Emotion Distribution Learning

Yuchang Dong and Xueqiang Zeng *

School of Computer & Information Engineering, Jiangxi Normal University, Ziyang Road 99,
Nanchang 330022, China; yixuxi_dong@jxnu.edu.cn
* Correspondence: xqzeng@jxnu.edu.cn

Abstract: Emotion distribution learning (EDL) handles emotion fuzziness by means of the emotion distribution, which is an emotion vector that quantitatively represents a set of emotion categories with their intensity of a given instance. Despite successful applications of EDL in many practical emotion analysis tasks, existing EDL methods have seldom considered the linguistic prior knowledge of affective words specific to the text mining task. To address the problem, this paper proposes a text emotion distribution learning model based on a lexicon-enhanced multi-task convolutional neural network (LMT-CNN) to jointly solve the tasks of text emotion distribution prediction and emotion label classification. The LMT-CNN model designs an end-to-end multi-module deep neural network to utilize both semantic information and linguistic knowledge. Specifically, the architecture of the LMT-CNN model consists of a semantic information module, an emotion knowledge module based on affective words, and a multi-task prediction module to predict emotion distributions and labels. Extensive comparative experiments on nine commonly used emotional text datasets showed that the proposed LMT-CNN model is superior to the compared EDL methods for both emotion distribution prediction and emotion recognition tasks.

1. Introduction

Emotion analysis aims to recognize and analyze the emotions behind massive human behavior data such as text, pictures, movies, and music [1–3]. With the rapid development of Internet-based social media, text-based emotion classification shows promising application prospects in many emerging artificial intelligence fields such as public opinion analysis, commodity recommendation, and business decision making [4]. In recent years, the text-oriented emotion classification model has become a research hotspot in the field of natural language processing and machine learning [5].

Many traditional emotion classification models adopt the multi-label learning paradigm and assume that an example is associated with some emotion labels. On this technical line, many scholars have proposed various effective works to solve the problem of emotion recognition. Multi-label learning can deal with multi-emotion recognition tasks, but it cannot quantitatively model a variety of emotions with different expression intensities [6]. To solve this problem, Zhou et al. proposed emotion distribution learning (EDL) [7] motived by label distribution learning (LDL) [6]. Different from the traditional emotion classification model, EDL associates an emotion distribution vector with each example (e.g., facial image or text sentence). The emotion distribution vector records the expression degree of a given example of each emotion label, and its dimension is the number of all emotions. In recent years, many EDL research works have been published in top conferences and journals in the field of machine learning. For example, Yang et al. proposed a circular-structured representation for visual emotion distribution learning by exploiting the intrinsic relationship between emotions based on psychological models [8]. Xu and Wang proposed a

method for learning emotion distribution based on an attention mechanism in 2021, using an emotional graph-based network to explore the correlation between various regions in the image and emotion distribution [9]. Fei et al. proposed a latent emotional memory network that can learn latent emotional distributions without external knowledge, and the model has been used well for classification tasks [10]. Zhao et al. proposed a small sample text EDL model of the meta-learning method [11]; Jia et al. proposed a facial EDL method using local low-rank label correlation [12]; Pang et al. proposed a basic model of an acceleration algorithm to predict the emotion distribution of unlabeled files [13]. These EDL methods can effectively record the intensity of examples of different emotion labels and show better performance than the traditional emotion classification model. However, most of the existing EDL methods do not introduce the unique affective word information containing prior emotion knowledge into the prediction model.

Affective words are words with different emotional tendencies [14], which have generally been manually labeled based on emotional linguistic knowledge. Different affective words are usually used to describe different emotional characteristics, and different combinations of affective words can also express different emotional tendencies. At present, some scholars have used affective word information in the field of emotion analysis. Teng et al. showed that affective words have a significant effect in predicting emotion in 2016 [14]. Zhang et al. proposed lexicon-based emotion distribution label enhancement (LLE) in 2018 [15]. Tong et al. annotated affective words manually and used them for emotion analysis [16]. These studies show that affective words can significantly improve the performance of the emotion analysis model. However, so far, there is no EDL method to use affective word information for emotion distribution prediction.

To address this problem, we propose a text emotion distribution learning model based on a lexicon-enhanced multi-task convolutional neural network (LMT-CNN). The overall architecture of the LMT-CNN model has three major modules: semantic information, emotion knowledge, and multi-task prediction. The semantic information module uses the sliding window convolution neural network to extract the semantic information of the text from the word embedding space of the input text. Based on the affective words extracted from the text, the emotion knowledge module uses the lexicon to introduce the corresponding emotional prior knowledge to synthesize an emotion knowledge vector. The input of the multi-task prediction module is constructed from the outputs of the first two modules. Then, the final emotion distribution is predicted through a full connection layer. The two prediction tasks of the emotion distribution output layer are emotion distribution prediction based on Kullback–Leibler (KL) loss [17] and emotion classification based on cross-entropy loss. The emotion with the highest score in the emotion distribution output layer is used as the dominant emotion output for emotion classification. The existing EDL research work shows that the multi-task convolution neural network model combined with KL loss and cross-entropy loss can achieve better performance by simultaneously training emotion distribution prediction and emotion classification tasks in an end-to-end manner [15].

Different from the existing EDL work based on neural networks, the proposed LMT-CNN method considers the linguistic prior knowledge of affective words unique to the text mining task and combines it with text-based semantic information to construct an end-to-end deep neural network. We evaluated the LMT-CNN method on the English emotion distribution dataset Semeval [18], Chinese emotion distribution dataset Ren-CECps [19], four English single-label emotion datasets (Fairy Tales [20], TEC [21], CBET, ISEAR [22], and Affect in Tweets [23]), and Chinese single-label emotion datasets NLP&CC 2013 and NLP&CC 2014 [24], comparing LMT-CNN methods with multiple baseline methods. Compared with the baseline methods, our LMT-CNN method achieved the best performance in almost all measurements.

The rest of the paper is organized as follows. First, Section 2 briefly reviews some related works. Then, Section 3 describes the proposed lexicon-enhanced multi-task convolutional neural network for emotion distribution learning. Section 4 describes in detail

the datasets used for the experiments, experimental setup, and experimental results and analysis. Section 5 mainly discusses the limitations of our proposed method. Finally, Section 6 concludes the entire paper and provides an outlook on future work.

2. Related Work

Emotions are ubiquitous in our daily life and play an important role in our lives, influencing our decisions and judgments. For a long time, people have expected a variety of ways to recognize the diverse emotions that humans generate. Many researchers have proposed a variety of algorithms aiming at making machines that have emotional intelligence through artificial intelligence algorithms, meaning machines that have the ability to recognize, interpret, and process emotions. For this reason, machines should first learn to recognize human emotions from their external and implicit emotional cues.

Emotion models are the basis of emotion recognition systems, which define the expression of emotions. To measure emotions quantitatively, psychologists considered that emotions exist in multiple states and therefore proposed various emotion models to distinguish different emotional states. Two of the most prominent emotion representation models are the categorical emotional state (CES) and the dimensional emotional space (DES) [25]. CES classifies emotions into different basic categories and considers each type of basic emotion independent. Among the popular CES models are binary emotions [26], Ekman's six emotions [27], and Plutchik's eight emotions [28]. Binary emotions contain positive, negative, and sometimes neutral emotions. In this case, "emotion" is often referred to as "sentiment". In the Paul Ekman model, emotions are independent and can be distinguished into six basic categories depending on how they are perceived by the experiencer. These basic emotions are joy, sadness, anger, disgust, surprise, and fear. Plutchik's eight emotions consist of amusement, anger, awe, contentment, disgust, excitement, fear, and sadness. The DES model assumes that emotions do not exist independently of each other and that there are interactions between emotions. Therefore, some scholars consider locating emotions in a multidimensional space, such as valence–arousal–dominance (VAD) [29]. In our work, Ekman's six emotions were used as the target emotion label set.

Early work in emotion recognition focused on combining machine learning methods to learn emotional features to recognize the emotions embedded in text, speech, or images. Examples include the methods of Naive Bayes, maximum entropy, and support vector machines [30]. Later, Vrysis et al. proposed a method for emotion recognition combined with lexicon-based and rule-based algorithms [31]. The lexicon-based approach relies on the semantic direction of the text with the polarity of the words and phrases appearing in it [32]. Rule-based algorithms design extraction rules based on syntactic dependencies [33], which need to be used within the controllable range of the rules and may lead to incorrect emotion judgments if they are beyond the range of the rules. We believe that the performance of this approach largely depends on the quality of the lexicon and the number and quality of the rules formulated. It is more suitable when the data are lacking, and there is a performance bottleneck when dealing with large-scale data, combined with the fact that now the path of the method based on deep learning language models is more attractive in the field of emotion classification [34].

Deep learning changes in feature engineering and feature learning both make problem solving easier. Traditionally, the efficiency of machine learning algorithms is highly dependent on how well the input data are represented. For this reason, feature engineering has been the most critical step in the machine learning workflow. In contrast, deep learning algorithms can automate feature extraction, which allows researchers to extract features with minimal domain knowledge and manpower. Another transformative aspect of deep learning is that models can learn all representation layers together at the same time. Through common feature learning, once the model modifies an internal feature, all other features that depend on that feature will automatically adapt accordingly without human intervention.

With the wide application of deep learning in the field of emotion recognition, various effective methods have been proposed by many scholars around the world. For example, Franzoni et al. proposed a semantic model for emotion recognition (SEMO), which can quickly extract the basic emotions hidden in short, emotion-rich sentences [35]. Dashtipour et al. proposed a novel context-aware multimodal framework for multimodal emotion analysis, which simultaneously uses text, speech, and visual cues to determine emotion [36]. Rahman et al. proposed a multimodal adaptation gate (MAG) as an attachment to BERT and XLNet which achieves human-level multimodal emotion analysis performance [37]. In recent years, emotion recognition methods have also been applied to social services [38] and opinion detection [39]. For example, Kydros et al. analyzed tweets from Twitter during the time of the coronavirus outbreak that helped government agencies develop effective strategies by understanding people's emotions [38]. Zou et al. designed an emotion analysis method that combines emojis and other emoticons with short texts to capture the deep value of public opinion and help companies optimize their services [40].

The traditional text emotion analysis model mainly classifies the polar emotion of the text, that is, to judge the positive and negative polarity of emotion. However, the polar emotion classification model cannot capture the fine-grained emotions contained in the text and can only be used in simple emotion analysis scenarios. Different from the traditional polar emotion analysis task, the goal of fine-grained emotion analysis is to identify fine-grained emotions in the text [41], such as anger, disgust, sadness, surprise, fear, and joy. The classical fine-grained emotion analysis model is generally modeled by single-label learning or multi-label learning. It is assumed that the example is associated with one or more emotion labels [6].

The classical multi-label learning model can handle many emotion recognition tasks, but its modeling ability is still insufficient to quantitatively answer the expression degree of each emotion label in the text [7]. In practical application, it is very common for a sentence to express multiple emotions at the same time. For example, the sentences in the commonly used Semeval text emotion dataset mark the expression degree of six fine-grained basic emotions [19]. As shown in Figure 1, fear is the main emotion of Figure 1a sentence, and the degree of expression is 53.2%. The expression of secondary emotion sadness is 39%. The fine-grained basic emotion labeling of Figure 1b sentence is similar to that of Figure 1a sentence.

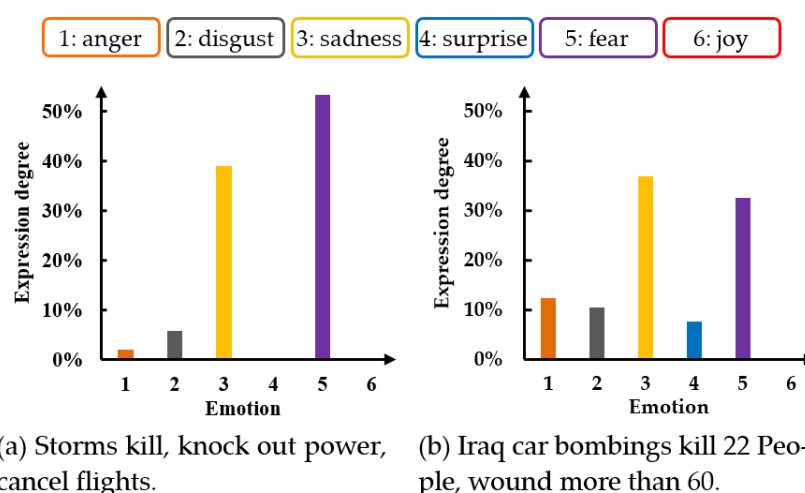**Figure 1.** Expression degree of two example sentences in the Semeval dataset of six basic emotions. (**a**) Emotion distribution of sentences with main emotion label is fear; (**b**) Emotion distribution of sentences with main emotion label is sadness.

Aiming to quantitatively deal with the situation that a sentence expresses different degrees of fine-grained emotions at the same time, Zhou et al. proposed emotion distribu-

tion learning (EDL) for the task of text emotion analysis [7]. EDL believes that the emotion of each sentence is a mixture of basic emotions with different expression intensities, and the goal of modeling is to transform the sentence s_i maps to an emotional distribution $\boldsymbol{D}_i = \left\{ d_{s_i}^{a_j} \right\}_{j=1}^{N}$, where $A = \{a_1, a_2, \ldots, a_N\}$ represents a finite set of emotion labels in a sentence, N is the number of emotion label classes, and $d_{s_i}^{a_j}$ is each element of emotional distribution that denotes sentence s_i expression degree of class j emotion label, $d_{s_i}^{a_j} \in [0, 1]$ and $\Sigma_j d_{s_i}^{a_j} = 1$. The EDL method assigns a distribution to each instance, and the strength of the probability distribution of each label indirectly indicates the relative importance of the corresponding label. To output emotion labels based on the emotion distribution, it is natural to consider one or more of the top-ranked emotions as the correct emotion label. Since the emotion classification datasets selected for the experiments in this paper are all multi-class datasets (each sample has only one emotion label in several possible classes), we assume that there is only one correct emotion label in a sample for simplicity of processing.

Emotion distribution learning can effectively deal with the problem that a sentence expresses multiple emotions with different intensities at the same time. It is more suitable for emotion analysis tasks with emotion fuzziness than the traditional emotion analysis model. In recent years, many scholars have put forward many effective works in the field of EDL [7,13,14,42–44]. For example, Zhou et al. proposed an EDL method based on the Plutchik's wheel of emotions in 2016 [7]; Zheng et al. proposed an EDL method using local correlation of samples in 2018 [42]; Batbaatar et al. proposed a semantic emotion neural network in 2019 [43]; Qin et al. proposed a constrained optimization method for cross-domain EDL in 2021 [44]. However, most of the existing EDL work only focuses on the semantic information or emotional wheel information, ignoring the prior knowledge of affective words in the text.

Affective words are words that express emotional tendencies in the text. Generally, they have been labeled with artificial emotion. The extraction and discrimination of affective words is the basic work of word-level emotion analysis and the basis of text emotion analysis. Relevant work has shown that emotion prior knowledge is very effective for emotion recognition tasks [14]. Scholars have proposed a variety of emotion recognition methods based on the lexicon. The general process of these methods is to extract affective words from the text based on the lexicon and then use the emotion labels of affective words to predict the emotion of the text. For example, Wang proposed a multi-constraint emotion classification model based on a lexicon in 2015 [45]; Zhang et al. proposed lexicon-based emotion distribution label enhancement (LLE) in 2018 [15]; Abdi et al. proposed a multi-feature fusion evaluation model based on deep learning for text emotion classification in 2019 [46]; Ke et al. proposed a new language expression model in 2020, which introduces part-of-speech tagging and affective word language knowledge into the training model [47]. These studies showed that using affective words in emotion recognition tasks can significantly improve the performance of the emotion analysis model, but most existing EDL methods do not consider the affective word information of a text. Compared with the existing EDL method, the LMT-CNN model proposed in this paper considers the prior knowledge of affective word linguistics specific to text mining tasks. According to the emotional category and number of affective words in the text, the emotional knowledge vector is generated, and the semantics information and emotional word knowledge are combined. KL loss and cross-entropy loss are combined to simultaneously learn emotion distribution prediction and emotion classification tasks in an end-to-end manner.

3. Text Emotion Distribution Learning Based on Lexicon-Enhanced Multi-Task Convolutional Neural Network

In order to improve the performance of the text emotion distribution learning model based on a deep network, this paper proposes a lexicon-enhanced multi-task convolutional neural network (LMT-CNN). The network structure of the LMT-CNN model includes

three modules: semantic information module, emotion knowledge module, and multi-task prediction module. The specific network structure is shown in Figure 2.
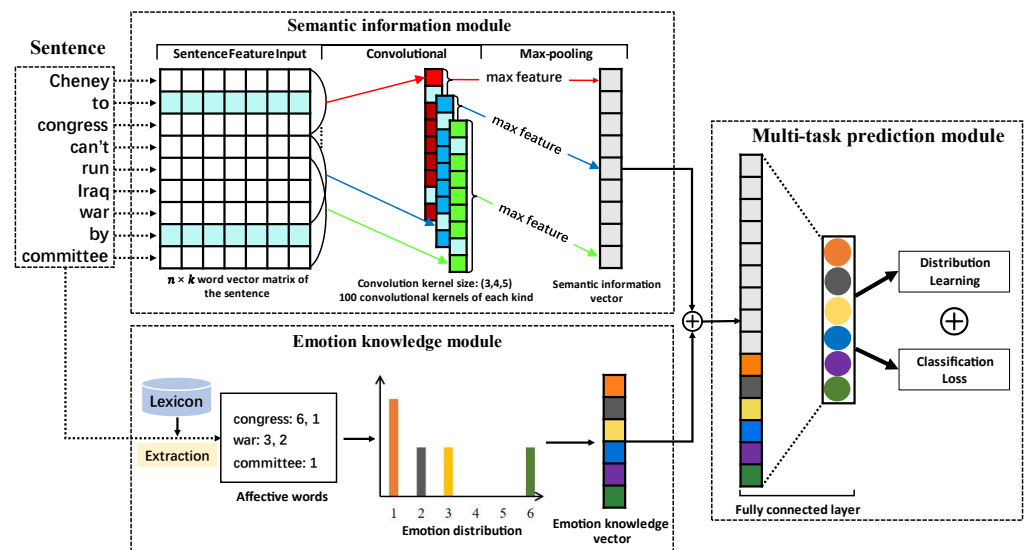

Figure 2. Overall framework of the LMT-CNN model (emotion labels 1: anger, 2: disgust, 3: sadness, 4: surprise, 5: fear, 6: joy).

3.1. Semantic Information Module

The semantic information module is a convolution neural network (CNN) built on word embedding for extracting semantic information from text. The CNN model was originally created in the domain of image recognition and classification, and its recent application in NLP has proven to be successful with excellent results [15]. Another competitive approach for semantic extraction is the long short-term memory (LSTM) network, which has been shown to have a good ability to learn sequential data [36]. LSTM controls the transmission state by gating the states, stores information that needs to be remembered for a long time, and forgets unimportant information. To the best of our knowledge, there is no evidence to prove which approach is definitely better for emotion recognition. However, because of the many contents introduced, LSTM leads to more parameters and makes the training more time-consuming. Since our aim was to verify the influence of prior knowledge of affective words on the model, we did not want to involve unnecessarily complex computations. Hence, we followed the previous successful EDL work by Zhang et al. [15] and used CNN to extract semantic information.

The semantic information module is composed of the input layer, convolution layer, and max-pooling layer. The specific workflow is as follows: Firstly, each sentence of the text dataset is used as the input of the module. Then, each word is transformed into a word embedding based on the pretrained word embedding model. Finally, the word embedding matrix representing the original sentence is processed by the convolution layer and max-pool layer to output a semantic information vector.

We use $S = \{(s_1, \boldsymbol{D}_1), (s_2, \boldsymbol{D}_2), \ldots, (s_n, \boldsymbol{D}_n)\}$, which represents the training text dataset, where n is the number of sentences in the dataset, s_i $(i \in \{1, 2, \ldots, n\})$ is the i-th sentence, $\boldsymbol{D}_i = \{d_{s_i}^{a_1}, d_{s_i}^{a_2}, \ldots, d_{s_i}^{a_N}\}$ represents for sentence s_i corresponding emotion distribution, $A = \{a_1, a_2, \ldots, a_N\}$ is a finite set of emotion labels, and N is the number of emotion labels. Next, we specifically describe the functions of the input layer, convolution layer, and max-pooling layer of this module.

Input Layer: The length of the input sentence is M, and $\boldsymbol{x}_i \in R^k$ is the k-dimensional word2vec word embedding representation of the i-th word in the sentence. In this way, the input sentence is represented as a $M \times k$ word embedding matrix $\boldsymbol{x}$ as

$$\boldsymbol{x} = \boldsymbol{x}_1 \oplus \boldsymbol{x}_2 \oplus \ldots \oplus \boldsymbol{x}_M, \tag{1}$$

where $\oplus$ is the concatenation operator. Note that, if the sentence length is less than M, the end of the word embedding matrix is filled with 0.

Convolution layer: A set of filters $w \in \mathbb{R}^{h \times k}$ with a sliding window size of h are adopted to generate new features. The width of the filter is the same as the width of the word embedding matrix, and thus the filter can only move in the height direction to conduct the convolution on several adjacent words. For example, set $x_{p:p+q}$ denotes the concatenation of word $x_p, x_{p+1}, \ldots, x_{p+q}$, and a new feature v_p is generated from the word window $x_{p:p+h-1}$ as

$$v_p = f\left(w \cdot x_{p:p+h-1} + b\right), \tag{2}$$

where $f(\cdot)$ is a nonlinear active function, and $b \in \mathbb{R}$ is a bias term. The filter is used for each possible word window in the sentence $x_{1:h}, x_{2:h+1}, \ldots, x_{M-h+1:M}$ to produce the feature mapping v by

$$v = [v_1, v_2, \ldots, v_{M-h+1}]. \tag{3}$$

Max-pooling layer: We use the standard max-pooling operation to obtain the generated feature map and take the maximum value of v as the characteristic of this particular filter:

$$v_{semantic} = \max(v), \tag{4}$$

this operation can capture the feature with the highest value as the most important feature for each feature map.

Through the above operations, a semantic feature can be extracted by each filter. The module uses several filters of different sizes to extract multiple semantic features from a given sentence.

3.2. Emotion Knowledge Module

The emotion knowledge module uses the lexicon to extract affective words from the sentence text and then synthesizes an emotion knowledge vector based on the prior emotion labels corresponding to affective words.

Given sentence s_i, we extract all affective words in s_i from the lexicon and obtain all the emotion labels corresponding to each emotion word. For each emotion, we set the number of emotion labels of corresponding affective words as m_{ij} and the total number of all emotion labels as C_i. Then, we set the emotion knowledge vector to $r_i = [r_{i1}, r_{i2}, \ldots, r_{iN}]$, where

$$r_{ij} = \begin{cases} \frac{m_{ij}}{C_i}, & if\ C_i \neq 0 \\ \frac{1}{N}, & else \end{cases}, j = 1, 2, \ldots, N. \tag{5}$$

Taking the input sentence in Figure 2 as an example, we extract three emotional words—congress, war, and committee—by looking up the lexicon. Among them, the affective word congress has two emotion labels (anger and joy), war has two emotion labels (disgust and sadness), and committee has one emotion label (anger). The total number of emotion labels of all affective words in the sentence is 5, the number of categories of emotional labels is 6, and the corresponding emotion knowledge vector is $\left[\frac{2}{5}, \frac{1}{5}, \frac{1}{5}, 0, 0, \frac{1}{5}\right]$.

In addition, when a sentence does not contain any affective words, the emotion knowledge weight of the sentence is the same on the six emotion labels, and its corresponding emotion knowledge vector is $\left[\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right]$.

3.3. Multi-Task Prediction Module

The multi-task prediction module supports two tasks: emotion distribution prediction and emotion classification, including the semantic synthesis layer and emotion distribution output layer.

Semantic synthesis layer: We splice the output $v_{semantic}$ of the semantic information module and the output $\boldsymbol{r}_i$ of the emotion knowledge module to generate the semantic synthesis vector $\hat{v}$ by

$$\hat{v} = v_{semantic} \oplus \alpha \cdot \boldsymbol{r}_i. \tag{6}$$

where $\oplus$ is the vector splicing operation, and the parameter α is used to control the ratio of adding emotional word information. When $\alpha = 0$, it means that the semantic synthesis layer only contains semantic information and does not consider the prior knowledge of affective words; when $\alpha = 1$, it means that all affective word information is considered in the semantic synthesis layer, and the semantic information is combined with all affective words prior to knowledge splicing. We believe that the importance of semantic information in the LMT-CNN model should be greater than affective word knowledge, that is, the value of parameters α should not be too large. This is because, in general, manually labeled emotion labels have high reliability, while the prior information of general affective words has more noise. This study analyzed the value of the parameter α in the experimental part where it is further discussed.

Emotion distribution output layer: Semantic synthesis vector $\hat{v}$ after a full connection layer transformation and the output result of the emotion distribution layer are obtained. The dimension of the emotion distribution layer is the number of categories of emotion labels.

The LMT-CNN model uses an end-to-end manner to train the emotion distribution prediction and emotion classification tasks simultaneously. For the distribution prediction task, we use KL loss [13] to measure the distance between the real distribution and the predicted one. The KL loss is defined as

$$E_{edl}(s, \boldsymbol{D}) = -\frac{1}{N} \sum_{i=1}^{d} \sum_{j=1}^{N} d_{s_i}^{j} \ln \ p_j^{(i)}, \tag{7}$$

where $p_j^{(i)} = \frac{e_j^{(i)}}{\sum_{t=1}^{N} e_t^{(i)}}$, $d_{s_i}^{j}$ represents the sum of each emotion label loss in sentence s_i.

The optimization objective of the emotion classification task is cross-entropy loss $E_{cls}(s, y)$, and its formula is

$$E_{cls}(s, y) = -\frac{1}{N} \sum_{i=1}^{d} \sum_{j=1}^{N} \mathbf{1}(y_i = j) ln p_j^{(i)}, \tag{8}$$

where $\mathbf{1}(y_i = j) = 1$ means when y_i is correctly divided into j classes; otherwise, $\mathbf{1}(y_i = j) = 0$.

The total loss of the emotion distribution output layer is a weighted combination of KL loss and cross-entropy loss as

$$E = (1 - \lambda) E_{cls}(s, y) + \lambda E_{edl}(s, \boldsymbol{D}), \tag{9}$$

where λ is a weight parameter to control the importance of two kinds of losses. According to the experimental results of the existing related work, we set $\lambda = 0.7$ [15].

Finally, we use the stochastic gradient descent (SGD) algorithm [48] to optimize the LMT-CNN model. When the prediction task is emotion classification, we take the emotion with the highest expression degree in the emotion distribution output from the full connection layer as the real emotion of the sentence.

Recent research work shows that the emotion distribution model based on a multi-task convolutional neural network (MT-CNN) [15] can achieve good results by simultaneously training emotion distribution prediction and emotion classification tasks in an end-to-end manner. However, the MT-CNN model does not consider the linguistic prior knowledge of affective words, which is very important in the task of text analysis. In the experimental part, the performance of LMT-CNN and MT-CNN models is compared and analyzed.

The source code of the LMT-CNN model is released at our website [49].

4. Experiments

To investigate the performance of the LMT-CNN model proposed in this paper, we conducted three groups of experiments on two emotion distribution datasets and five single-label emotion datasets. First, to verify the effectiveness of adding affective word information to the LMT-CNN model, we changed the value of α and recorded the performance of LMT-CNN. Second, to test the performance of the emotion distribution prediction of the LMT-CNN model, we used 8 different emotion distribution learning methods compared with the LMT-CNN method. Third, to examine the performance of the LMT-CNN model on emotion classification tasks, we used the deep-learning-based TextCNN, MT-CNN, and LMT-CNN emotion distribution learning models for comparative experiments.

4.1. Dataset

There were 9 commonly used text emotional datasets adopted in the experiments, namely Semeval [18], Ren-CECps [19], Fair Tales [20], TEC [21], CBET, ISEAR [22], Affect in Tweets [23], NLP&CC 2013, and NLP&CC 2014 [24]. Among them, Semeval and Ren-CECps are emotion distribution datasets, and the other 7 datasets are traditional single-label datasets. We list the details of all experimental datasets in Table 1.

Table 1. Experimental datasets (the sentence number of each emotion in the SemEval and Ren-CECps datasets is the sentence number whose dominant emotion is the corresponding emotion).

Dataset	Anger	Disgust	Sadness	Surprise	Fear	Joy	All
SemEval	91	42	192	435	262	228	1250
Ren-CECps	1117	2293	5841	536	-	4137	13,924
Fairy Tales	216	-	264	114	166	444	1204
TEC	1555	761	3830	3849	2816	8240	21,051
CBET	8540	8540	8540	8540	8540	8540	51,240
ISEAR	1087	1081	1083	-	1090	1090	5431
Affect in Tweets	844	-	747	-	1105	793	3489
NLP&CC 2013	1146	2075	1562	482	186	2140	7581
NLP&CC 2014	1899	3130	2478	820	299	2805	11,431

SemEval is an English text dataset containing multiple emotions, including 1250 news headlines labeled with 6 emotion labels (anger, disgust, sadness, surprise, fear, and joy). Each headline has a percentile score for each emotion. The multiple emotions associated with each sample can be regarded as an emotion distribution vector, and the length of the vector is 1 through normalization.

Ren-CECps is a Chinese text emotion distribution dataset. The sentence corpus comes from Chinese blogs. Each sentence is labeled with 8 emotion labels, and each sentence has a score for each emotion. In order to be consistent with the types of emotions in other datasets, we selected 13,924 sentences of 5 emotions (anger, disgust, sadness, surprise, and joy) for experiments. The emotion distribution label retains the scores of the corresponding five emotions, and the multiple emotion labels corresponding to each sentence can be regarded as an emotion distribution vector, and the length of the vector is 1 by normalization.

Fairy Tales contains 1204 sentences from 185 children's stories, each sentence labeled with one of five emotions (anger, sadness, surprise, fear, and joy). TEC contains 21,051 emotional tweets, each marked with one of six emotions (anger, disgust, sadness, surprise, fear, and joy). CBET is marked with 76,860 tweets of 9 emotions, and 8540 tweets are collected for each emotion. We retained a total of 51,240 tweets of 6 emotions. ISEAR contains 7666 English sentences. Each English sentence describes the situation and experience of different people when they experience seven main emotions. We retained five emotions, a total of 5431 sentences. Affect in Tweets is created from tweets. There are four emotions contained in dataset: anger, fear, joy, and sadness. The dataset contains 3489 English comments from tweets.

NLP&CC 2013 is a Chinese Weibo corpus containing 7 emotion labels released by the Natural Language Processing and Chinese Computing Conference in 2013, containing 10,552 emotional sentences and 29,633 non-emotional sentences. In this experiment, 7581 emotional sentences marked with six emotions of anger, disgust, sadness, surprise, fear, and joy were retained. Table 1 respectively lists the detailed information of the dataset. NLP&CC 2014 Chinese microblog emotion analysis dataset is from Sina Weibo, and all tweets in the dataset are divided into two categories with or without emotion; those without emotion are labeled as none, and those with emotion are divided into 7 emotion categories. In this paper, six categories of emotions were selected: happiness, anger, sadness, fear, disgust, and surprise; a total of 11,431 Chinese tweets were selected.

The English lexicon used in the experiment was formed by the merger of Emosenticnet [50] and NRC [51]. The Emosenticnet lexicon contains 13,189 English affective words, and each affective word is labeled with 6 emotion labels. Each affective word in Emosenticnet and NRC lexicons is marked with one or more emotion labels. When the two lexicons were merged, we retained the 6 emotion labels (anger, disgust, sadness, surprise, fear, and joy) and deleted the affective words that were not labeled with these 6 emotions. For an affective word shared by the two dictionaries, the emotion label is the union of the emotion label of the affective word in the two lexicons. In the end, we obtained an English affective lexicon consisting of 15,603 affective words and 6 emotion labels. The Chinese affective lexicon uses the emotion vocabulary text database of the Dalian University of Technology [52], which has 27,466 affective words and 11 emotion labels. We kept 5 emotion labels (anger, disgust, sadness, surprise, and joy) corresponding to the Chinese dataset and deleted the affective words that were not labeled with these 5 emotions. Finally, 15,179 Chinese affective words were retained, and each word was labeled with 1 label.

4.2. Experimental Setup

The experimental process adopted the standard stratified 10-fold cross-validation. We partitioned the dataset before feeding the data into the network. All samples were divided into ten categories evenly under the constraint of keeping the category proportions essentially the same. Each time, one of the samples was selected as the test set, and the remaining data were merged as the training dataset. In total, 90% of the training set was used for neural network training, and the remaining 10% was used as the validation set. Each fold cross-validation was a separate emotion prediction task, repeated ten times, and the average evaluation indicators of the ten cross-validations were used to evaluate the final performance of the model. The data division of all models participating in the comparison experiment remained the same.

Since the output of the LDL model is a distribution vector, a single indicator can reflect only one aspect of the algorithm on a particular data, and it is difficult for us to determine which indicator is the best. Geng et al. proposed that when comparing different LDL algorithms in the same dataset, multiple indicators can be used to evaluate and compare the algorithms [6]. Following four principles, six indicators were finally selected to evaluate the LDL algorithms. Based on this suggestion, Zhou et al. chose six indicators to measure the average similarity or distance between the actual and predicted emotion distributions in the EDL model, respectively [7]. Similarly, we used six evaluation indicators for our emotion distribution prediction task evaluation, namely *Euclidean, Sørensen, Squaredχ^2, KL Divergence, Cosine,* and *Intersection* [7]. The calculation formulas of the 6 evaluation indicators are as follows:

$$Euclidean(P,Q) = \sqrt{\sum_{i=1}^{N}(P_i - Q_i)^2} \tag{10}$$

$$Sorensen(P,Q) = \frac{\sum_{i=1}^{N}|P_i - Q_i|}{\sum_{i=1}^{N}|P_i + Q_i|} \tag{11}$$

$$Squared\chi^2(P,Q) = \sum_{i=1}^{N}\frac{(P_i - Q_i)^2}{P_i + Q_i} \tag{12}$$

$$K - L_{(P,Q)} = \sum_{i=1}^{N} P_i ln \frac{P_i}{Q_i} \tag{13}$$

$$Cosine(P,Q) = \frac{\sum_{i=1}^{N}(P_i \times Q_i)}{\sqrt{\sum_{i=1}^{N} P_i{}^2} \times \sqrt{\sum_{i=1}^{C} Q_i{}^2}} \tag{14}$$

$$Intersection(P,Q) = \sum_{i=1}^{N} min(P_i, Q_i) \tag{15}$$

where P is the true emotion distribution of the text, Q is the predicted emotion distribution, and N is the number of emotion types contained in the sentence. These 6 indicators are used to measure the similarity between the predicted emotion distribution Q and the real emotion distribution P.

The emotion classification task uses 4 commonly used classification task evaluation indicators, namely *Precision, Recall, F1-score* and *Accuracy*. The specific calculation formulas of the 4 evaluation indicators are as follows:

$$Precision = \frac{|TP|}{|TP + FP|} \tag{16}$$

$$Recall = \frac{|TP|}{|TP + FN|} \tag{17}$$

$$F1 - score = \frac{2 \times precision \times recall}{precision + recall} \tag{18}$$

$$Accuracy = \frac{|TP + TN|}{|A|} \tag{19}$$

where *TP* represents the number of positive samples with correct predictions, *FP* represents the number of negative samples with incorrect predictions, *FN* represents the number of positive samples with incorrect predictions, *TN* represents the number of negative samples with correct predictions, and A represents the total number of all samples. *Precision* represents the proportion of all samples predicted to be positive that are actually positive. *Recall* indicates the proportion of all positive examples in the sample that are correctly predicted. The *F1-score* is the weighted average (or harmonic mean) of *Precision* and *Recall*. Therefore, this score takes both false positives and false negatives into account to strike a balance between *Precision* and *Recall*. *Accuracy* is the most intuitive measure of performance and reflects the proportion of samples correctly classified by the prediction model.

For all the datasets we used, we set the filter windows of CNN to 3, 4, 5 in our experiments, each with 100 feature maps; the dropout rate was 0.5, the learning rate was set to 0.05, and the mini-batch size was 50 [53]. We used the word embedding method to convert each instance into a word embedding matrix. We counted the average length of Chinese dataset samples as 100 and the average length of English dataset samples as 15, and thus we set the word vector matrix shape as 100×300 and 15×300, respectively. We used publicly available word2vec vectors. The English word vector is trained from 100 billion words in Google News [54], and the Chinese word vector is trained from Sogou News [55]. The dimension of all these word vectors is 300. In particular, the parameters of the LMT-CNN model were set as shown in Table 2.

For the computational complexity of the LMT-CNN model, we chose Params and FLOPs to measure the complexity. Params refers to the total number of parameters to be trained in the model training and is only related to the defined network structure. FLOPs is the number of floating-point operations, which can be regarded as the calculation amount of the model [56]. We regard the multi-add combination as a floating-point operation. FLOPs are related to different layer operation structures, and the maximum length of the input sentences influences the size of the computation. We calculated the parameter quantities of the convolutional layer and the fully connected layer that use different height convolution kernels; we also calculated the calculation amount of convolutional layer and

fully connected layer at different sentence maximum length *M*. The results are shown in Table 3.

Table 2. LMT-CNN model parameter settings.

Parameter	Set Value
Word vector dimension	300
Sliding window	(3, 4, 5)
Optimizer	SGD
Learning rate	0.05
Dropout	0.5
Batch size	50
Word embedding matrix dimension in Chinese text	100×300
Word embedding matrix dimension in English text	15×300

Table 3. The computational complexity of the LMT-CNN model.

Layers	Params	MFLOPs	
		M = 15	*M* = 100
Conv2d (k_h = 3)	90,100	234.260	1765.960
Conv2d (k_h = 4)	120,100	312.260	2353.960
Conv2d (k_h = 5)	150,100	390.260	2941.960
Fully connected	1842	0.004	0.004
Total	362,142	936.784	7061.884

4.3. The Influence of the Weight Coefficient α of Affective Words on the Performance of LMT-CNN Model

The affective word weight coefficient *α* is an important parameter of the LMT-CNN model, which is used to control the weight of affective word information and semantic information. To verify the effectiveness of adding affective word information to the LMT-CNN model, we changed the value of *α* from 0 to 1.5 (every 0.1) and recorded the scores of the *Euclidean*, *KL Divergence*, *Cosine*, and *Accuracy* indicators of the LMT-CNN model on the Semeval and Ren-CECps datasets. The detailed experimental results are shown in Figures 3 and 4.

(a) (b)

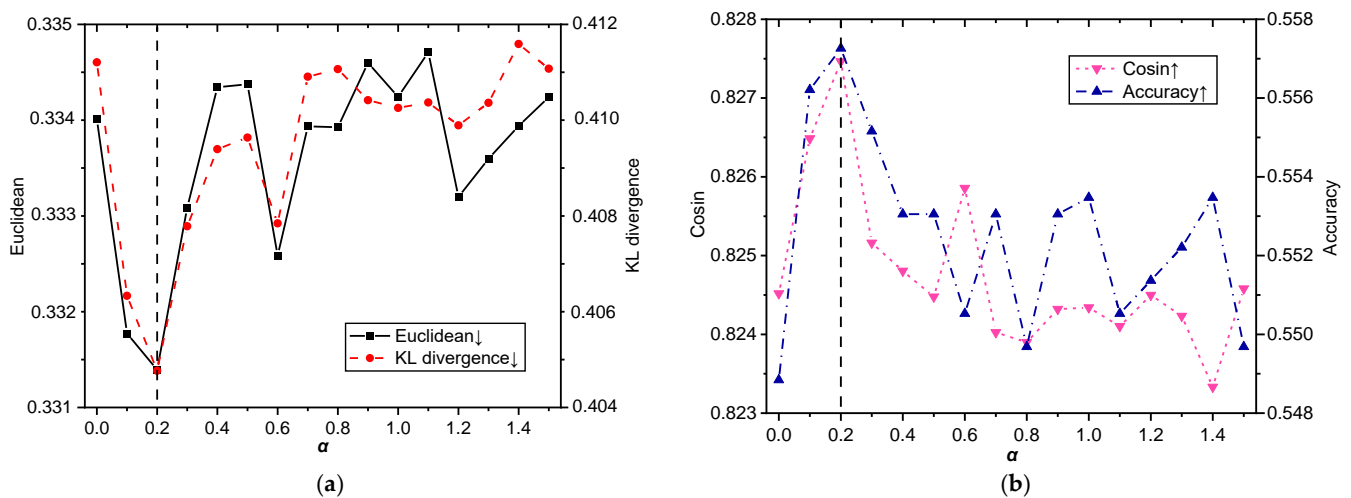**Figure 3.** The influence of the weight coefficient *α* on the performances of the LMT-CNN model on the Semeval dataset (↑ means that the larger the indicator is, the better, and ↓ means that the smaller the indicator is, the better). (**a**) *Euclidean* and *KL Divergence*; (**b**) *Cosine* and *Accuracy*.

Figure 4. The Influence of the weight coefficient α on the performances of the LMT-CNN model on the Ren-CECps dataset ($\uparrow$ means that the larger the indicator is, the better, and $\downarrow$ means that the smaller the indicator is, the better). (**a**) *Euclidean* and *KL Divergence*; (**b**) *Cosine* and *Accuracy*.
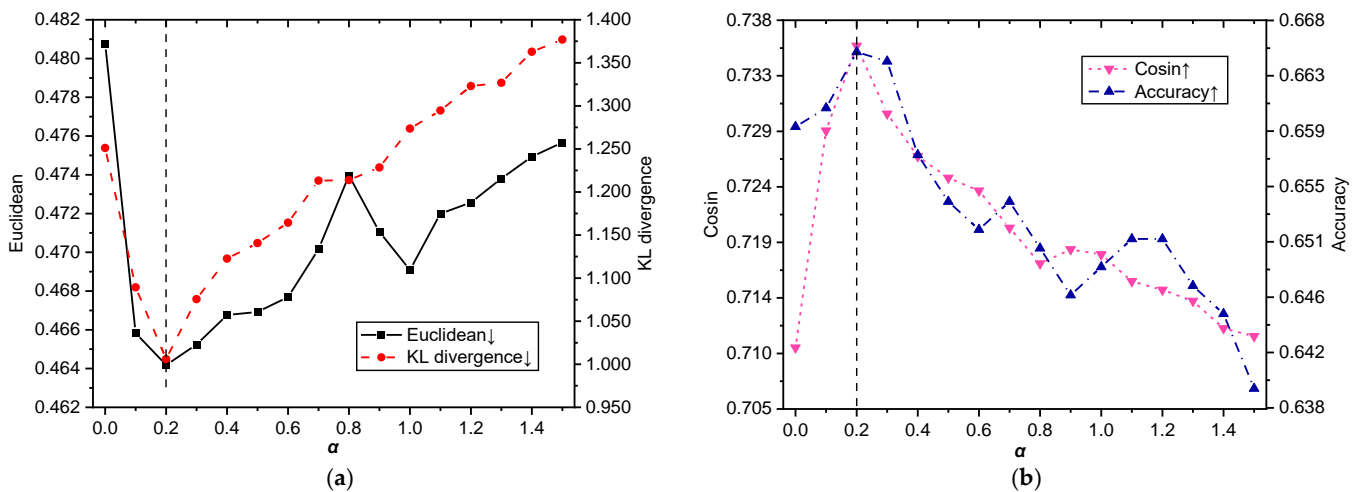
As can be seen from the experimental results of the English dataset in Figure 3, although the scores of the four evaluation indicators vary greatly with the change in weight α, the optimal value is reached when $\alpha = 0.2$. When α increases from 0 to 0.2, the two indicators in Figure 3a gradually decrease, and the two indicators in Figure 3b gradually increase, which shows that it is beneficial to increase the linguistic knowledge of affective words. When $\alpha = 0.2$, the four indicators in Figure 3 reach the best value, and the performance of the LMT-CNN model is the best, which shows that emotion knowledge and semantic information reach a balance. When $\alpha > 0.2$, the indicators in Figure 3a,b show an unstable upward and downward trend, respectively, and the performance of the model begins to decrease as a whole, which shows that when the weight of affective words is too large, the noise of affective word information affects the performance of the model.

For the Chinese dataset Ren-CECps, it can be seen from the experimental results in Figure 4 that the changing trend of the four evaluation indicators with the increase in weight is similar to that of the English dataset. The *Euclidean* and *KL Divergence* indicators in Figure 4a rise first and then fall; *Cosine* and *Accuracy* indicators in Figure 4b fall first and then rise. The four evaluation indicators are in $\alpha = 0.2$, indicating that the affective word information and semantic information reach a balance.

The above experimental results show that affective words contain effective emotional information. Adding affective word information to the EDL model can help improve the performance of emotion distribution prediction. However, since affective words have more noise, affective word information weight α should not be too large. Considering the balance of text semantic information and affective word information, in the following experiments of this paper, we set $\alpha = 0.2$ for the dataset for the LMT-CNN model.

4.4. Contrast Experiment of Multiple Text Emotion Distribution Prediction Methods

To verify the performance of the emotion distribution prediction of the LMT-CNN model, we selected 8 different EDL methods as baseline methods and conducted a comprehensive comparative experiment on the Semeval and Ren-CECps datasets. The compared EDL methods included AA-KNN, AA-BP, SA-LDSVR, SA-IIS, SA-BFGS, SA-CPNN [6], TextCNN [53], and MT-CNN [15]. The detailed model configurations are described as follows.

Specifically, AA-KNN and AA-BP are extended versions of the classical KNN algorithm and BP (backpropagation) neural network to solve LDL tasks [6]. In the AA-KNN model, we set the number of neighbor samples as 4, used *Euclidean* distance to calculate the distance from the sample point to be classified to each other sample, and predicted

the category of the sample to be tested through the nearest 4 neighboring samples. In the AA-BP model, we created a three-layer network; the number of neurons in the first layer was 24, and the transfer function was "*tansig*". The number of neurons in the second layer was 60, and the transfer function was "*tansig*". The number of neurons in the output layer was 6, and the transfer function was "*purelin*". We set the learning rate to 0.05 and the momentum factor to the default value of 0.9.

Additionally, SA-LDSVR, SA-IIS, SA-BFGS, and SA-CPNN are algorithms specifically designed for LDL tasks [6]. In the SA-LDSVR model, we chose radial basis function (RBF) as the core function, the penalty parameter was 1.0, and the other parameters were the default values. The SA-IIS method establishes the maximum entropy model and uses the IIS algorithm to estimate the parameters. SA-BFGS follows the idea of an effective quasi-Newton method BFGS to further improve IIS-LLD. SA-CPNN has a network structure similar to Modha's neural network, where the only difference is its training manner is supervised.

Both TextCNN and MT-CNN are deep-neural-network-based models. TextCNN is a convolutional neural network model for text emotion classification [53]. The height of convolutional kernel size was divided into three groups (3, 4, 5), and the width was 300, which was equal to the dimension of the word vectors. There were 32 channels in each group. Batch size and learning rate were set to 16 and 0.001. MT-CNN is a multi-task convolutional neural network model that simultaneously predicts the distribution of text emotion and the dominant emotion of the text [15]. The height of convolutional kernel size was divided into three groups (3, 4, 5), and the width was 300, which was equal to the dimension of the word vectors. We combined the cross-entropy loss with the KL loss by setting the weight (λ). Therefore, considering the balance between distribution prediction and classification performance in the experiment, $\lambda = 0.7$ [15] was set for this model, which means that the weight of the cross-entropy loss was 0.3, and the weight of the KL loss was 0.7.

The detailed experimental results of all emotion distribution learning methods participating in the comparison of the datasets are shown in Tables 4 and 5. In the table, ↑ means that the larger the indicator is, the better, and ↓ means that the smaller the indicator is, the better. The optimal results of each indicator are marked in bold.

Table 4. Experimental results comparing the performance of 9 emotion distribution learning methods for emotion distribution prediction and emotion classification on the SemEval dataset (↑ indicates that the larger the indicator is, the better; ↓ indicates that the smaller the indicator is, the better.). The best performances of each indicator are marked in bold.

EDL Method	Emotion Distribution Prediction						Emotion Classification			
	Euc (↓)	Sør (↓)	Squ (↓)	KL (↓)	Cos (↑)	Int (↑)	Pre (%)	Rec (%)	F1 (%)	Acc (%)
AA-KNN	0.5249	0.4883	0.5649	0.8729	0.6064	0.5117	26.14	19.52	15.33	36.80
AA-BP	0.4677	0.4483	0.4775	0.7858	0.6915	0.5517	38.83	32.13	33.32	43.20
SA-LDSVR	0.4489	0.4360	0.4262	0.5982	0.7367	0.5640	23.77	25.87	21.16	37.60
SA-IIS	0.4804	0.4719	0.4922	0.6871	0.6954	0.5281	26.50	24.08	21.99	39.20
SA-BFGS	0.5573	0.4839	0.5245	1.4149	0.6285	0.5161	32.22	31.55	31.41	40.00
SA-CPNN	0.6688	0.6003	0.7849	2.3375	0.4780	0.3997	21.94	19.25	19.53	23.20
TextCNN	0.7100	0.6068	0.9746	2.0047	0.5353	0.3932	37.29	32.01	33.16	42.65
MT-CNN	0.3344	0.3200	0.3636	0.4118	0.8239	0.6800	44.81	**42.05**	41.27	54.96
LMT-CNN	**0.3324**	**0.3177**	**0.3586**	**0.4057**	**0.8261**	**0.6823**	**45.20**	42.00	**41.68**	**55.76**

Table 5. Experimental results comparing the performance of 9 emotion distribution learning methods for emotion distribution prediction and emotion classification on the Ren-CECps dataset (↑ indicates that the larger the indicator is, the better; ↓ indicates that the smaller the indicator is, the better). The best performances of each indicator are marked in bold.

EDL Method	Emotion Distribution Prediction						Emotion Classification			
	Euc (↓)	Sør (↓)	Squ (↓)	KL (↓)	Cos (↑)	Int (↑)	Pre (%)	Rec (%)	F1 (%)	Acc (%)
AA-KNN	0.7918	0.7826	1.1809	**0.7575**	0.3546	0.2115	12.53	12.43	11.44	16.47
AA-BP	0.7042	0.7465	0.9998	1.3841	0.5049	0.2535	8.37	12.42	5.76	24.83
SA-LDSVR	0.7050	0.7579	1.0166	1.3969	0.5006	0.2421	8.79	12.55	7.39	24.00
SA-IIS	0.7073	0.7546	1.0132	1.4031	0.4967	0.2454	12.16	12.65	8.97	22.80
SA-BFGS	0.7079	0.7495	1.0069	1.4081	0.4966	0.2504	12.01	.12.93	9.91	23.08
SA-CPNN	0.7450	0.8015	1.1136	1.6538	0.4159	0.1984	9.88	12.41	4.96	19.14
TextCNN	0.5033	0.3945	0.6410	1.0831	0.7043	0.6055	54.65	47.32	48.86	62.57
MT-CNN	0.4990	0.3883	0.6293	1.0629	0.7076	0.6117	56.07	47.54	49.71	63.72
LMT-CNN	**0.4682**	**0.3657**	**0.5865**	1.0309	**0.7330**	**0.6343**	**58.72**	**50.27**	**52.22**	**66.95**

In the emotion classification task on Semeval dataset, except that MT-CNN *Recall* is higher than LMT-CNN, LMT-CNN model is 0.39%, 0.41%, and 0.80% higher than MT-CNN model in *Precision*, *F1-score*, and *Accuracy*. On Ren-CECps dataset, LMT-CNN is 2.65%, 3.73%, 2.51%, and 3.23% higher than MT-CNN in *Precision*, *Recall*, *F1-score*, and *Accuracy*. The experiment results show that the emotion knowledge vector extracted from the emotion lexicon can effectively increase the dominant emotion information of sentences and help to improve the emotion classification performance of the LMT-CNN model.

4.5. Comparison of Emotion Classification Performance on Single-Label Datasets

To examine the performance of the LMT-CNN model on the emotion classification task, we conducted comparative experiments using deep-learning-based TextCNN, MT-CNN, and LMT-CNN emotion distribution learning models on seven single-label datasets. Note that the LLE label enhancement method [12] was used to transform the single-label dataset into the emotion distribution dataset.

When the classification task is multi-category, the performance of the classification model can be evaluated by comparing the scores of the classification indicators for each individual category. If using the arithmetic mean of all categories on different indicators as a criterion for evaluating the model, in the case of data imbalance, the categories with fewer data will affect the scores of the indicators more. For example, the *Precision* indicator may score relatively high on categories with small sample sizes. This improves the average of *Precision* indicators on the overall data to some extent, while in reality, not so many samples are correctly classified. Therefore, we compared the performance of the three CNN-based EDL models on each emotion label. The detailed comparative experimental results for each individual emotion category are shown in Table 6, where the last column shows the macro averaged score of the corresponding indicators. The optimal results of each indicator are marked in bold.

As can be seen from Table 6, LMT-CNN achieves the overall best results on all seven datasets. Taking the macro averaged *F1-score* as an example, on the Fairy Tales, TEC, CBET, ISEAR, and Affect in Tweets datasets, the macro averaged F1 value of the LMT-CNN model is 0.40%, 1.32%, 0.91%, 2.17%, and 3.36% higher than that of the MT-CNN model and 0.64%, 2.37%, 1.88%, 3.54%, and 4.78% higher than that of TextCNN. Comparing the performance of LMT-CNN and the baseline model on each emotion, we can see that our model gives better results on almost all emotions. Moreover, the classification of Anger, Sadness and Fear emotions is more challenging, and the performance in all models is relatively low. For example, in the experimental results of Chinese datasets NLP&CC 2013 and NLP&CC 2014, the scores of *Recall* and *F1-score* on the three emotion labels Anger, Sadness, and Fear were relatively low, while the classification on the two labels Joy and Surprise achieved better results.

Compared to the two baseline methods, LMT-CNN shows different degrees of performance score improvement on different datasets. Overall, on all seven datasets, the LMT-CNN method almost always scored higher than the baseline method being compared for the classification indicators in each class. This shows that the classification performance is improved more effectively by combining the emotion knowledge module with semantic connotations for the same input text features. Adding the consideration of prior knowledge of affective words clearly helps in emotion recognition, and our particular method learns more information in the sentence while considering the information of affective words.

In addition, consistent with the experimental results of Zhang et al. [15], the MT-CNN model scored higher than the TextCNN model on each evaluation indicator in the experiments on the seven datasets. This result shows that the performance of the multi-task neural network is significantly better than that of the traditional single-task networks in the emotion classification task. MT-CNN and LMT-CNN use the combination of cross-entropy loss and KL loss to train multi-task neural networks. In the training process, emotion distribution prediction and emotion classification tasks improve each other and can achieve better performance than single-task CNN.

Table 6. Comparative performance of three CNN-based EDL models for emotion classification. The best performances of each indicator on each dataset are marked in bold.

Dataset	Model	Indicator	Emotion						Averaged
			Anger	Disgust	Sadness	Surprise	Fear	Joy	
Fairy Tales	TextCNN	Pre (%)	82.05%	71.24%	66.05%	-	66.15%	76.06%	72.31%
		Rec (%)	65.53%	76.46%	71.20%	-	65.58%	73.55%	70.46%
		F1 (%)	63.14%	68.35%	68.87%	-	72.87%	76.39%	69.92%
	MT-CNN	Pre (%)	82.13%	71.51%	66.17%	-	66.85%	76.56%	72.64%
		Rec (%)	65.83%	76.67%	71.23%	-	66.10%	73.48%	70.66%
		F1 (%)	63.75%	68.35%	69.42%	-	73.17%	76.09%	70.16%
	LMT-CNN	Pre (%)	**82.31%**	**71.67%**	**66.67%**	-	**67.14%**	**76.92%**	**72.94%**
		Rec (%)	**66.96%**	**77.29%**	**71.50%**	-	**66.32%**	**73.77%**	**71.17%**
		F1 (%)	**64.48%**	**68.89%**	**69.80%**	-	**73.23%**	**76.42%**	**70.56%**
TEC	TextCNN	Pre (%)	52.38%	48.74%	49.13%	57.12%	65.70%	55.66%	54.79%
		Rec (%)	31.67%	30.74%	76.57%	43.57%	48.49%	52.91%	47.33%
		F1 (%)	38.36%	54.05%	72.25%	46.04%	30.93%	57.33%	49.83%
	MT-CNN	Pre (%)	55.59%	49.44%	51.57%	58.51%	67.56%	55.56%	56.37%
		Rec (%)	30.60%	30.97%	78.56%	44.87%	48.60%	53.85%	47.91%
		F1 (%)	38.47%	53.05%	72.98%	45.17%	35.77%	59.86%	50.88%
	LMT-CNN	Pre (%)	**56.36%**	**50.38%**	**52.74%**	**58.77%**	**68.33%**	**55.78%**	**57.06%**
		Rec (%)	**35.30%**	**30.81%**	**78.90%**	**45.73%**	**50.31%**	**58.79%**	**49.97%**
		F1 (%)	**38.99%**	**55.00%**	**73.08%**	**48.45%**	**37.65%**	**60.04%**	**52.20%**
CBET	TextCNN	Pre (%)	57.82%	54.78%	53.19%	57.67%	63.87%	66.80%	59.02%
		Rec (%)	50.63%	61.57%	61.90%	52.21%	63.53%	72.48%	60.39%
		F1 (%)	54.56%	59.09%	68.50%	50.03%	62.68%	70.89%	60.96%
	MT-CNN	Pre (%)	59.01%	57.70%	53.43%	59.77%	73.19%	69.39%	62.08%
		Rec (%)	51.73%	64.23%	67.22%	53.63%	64.35%	72.45%	62.27%
		F1 (%)	54.66%	60.10%	69.29%	50.88%	65.60%	71.06%	61.93%
	LMT-CNN	Pre (%)	**62.09%**	**60.40%**	**54.78%**	**62.80%**	**75.19%**	**67.81%**	**63.85%**
		Rec (%)	**53.09%**	**65.51%**	**68.32%**	**54.58%**	**65.44%**	**72.90%**	**63.31%**
		F1 (%)	**55.92%**	**60.42%**	**70.74%**	**51.46%**	**66.60%**	**71.91%**	**62.84%**

Table 6. *Cont.*

| Dataset | Model | Indicator | Emotion | | | | | | Averaged |
			Anger	Disgust	Sadness	Surprise	Fear	Joy	
ISEAR	TextCNN	Pre (%)	61.36%	63.50%	76.64%	-	70.67%	79.30%	70.29%
		Rec (%)	70.84%	64.24%	74.21%	-	71.66%	64.59%	69.11%
		F1 (%)	62.14%	65.22%	76.39%	-	72.09%	73.97%	69.96%
	MT-CNN	Pre (%)	61.31%	64.68%	80.27%	-	72.16%	81.13%	71.91%
		Rec (%)	71.62%	64.46%	77.37%	-	73.66%	69.36%	71.29%
		F1 (%)	65.68%	67.63%	77.00%	-	74.25%	72.09%	71.33%
	LMT-CNN	Pre (%)	**62.28%**	**66.00%**	**82.07%**	-	**72.50%**	**82.15%**	**73.00%**
		Rec (%)	**72.38%**	65.10%	79.34%	-	**74.40%**	**71.64%**	**72.57%**
		F1 (%)	**66.54%**	**70.64%**	80.68%	-	**74.95%**	**74.69%**	**73.50%**
Affect in Tweets	TextCNN	Pre (%)	60.60%	55.26%	-	-	34.69%	28.57%	44.78%
		Rec (%)	43.59%	39.60%	-	-	53.12%	21.11%	39.36%
		F1 (%)	41.02%	49.12%	-	-	40.98%	29.05%	40.04%
	MT-CNN	Pre (%)	60.83%	55.57%	-	-	33.57%	32.67%	45.66%
		Rec (%)	49.33%	40.18%	-	-	53.42%	24.29%	41.80%
		F1 (%)	43.97%	49.38%	-	-	43.04%	29.05%	41.36%
	LMT-CNN	Pre (%)	62.71%	56.18%	-	-	35.39%	36.51%	47.70%
		Rec (%)	50.70%	41.21%	-	-	53.61%	25.26%	42.70%
		F1 (%)	45.22%	50.07%	-	-	44.25%	30.00%	42.39%
NLP&CC 2013	TextCNN	Pre (%)	44.86%	49.12%	52.37%	58.72%	50.00%	73.03%	54.68%
		Rec (%)	32.97%	42.08%	30.59%	52.59%	38.38%	67.00%	43.93%
		F1 (%)	34.84%	55.38%	36.32%	47.15%	33.76%	71.83%	46.55%
	MT-CNN	Pre (%)	46.55%	49.77%	57.06%	63.27%	53.11%	78.07%	57.97%
		Rec (%)	34.01%	48.26%	32.86%	53.92%	39.26%	**69.86%**	46.36%
		F1 (%)	37.94%	53.98%	36.65%	51.38%	37.82%	73.45%	48.54%
	LMT-CNN	Pre (%)	**51.55%**	**51.18%**	**60.00%**	**66.97%**	**54.03%**	**78.91%**	**60.44%**
		Rec (%)	**36.56%**	**50.90%**	**32.68%**	**58.13%**	**38.68%**	69.80%	**47.79%**
		F1 (%)	**38.45%**	**56.38%**	**38.82%**	**52.82%**	**38.78%**	**73.55%**	**49.80%**
NLP&CC 2014	TextCNN	Pre (%)	40.60%	45.03%	62.20%	59.63%	51.11%	70.04%	54.77%
		Rec (%)	36.36%	60.14%	25.18%	58.42%	24.39%	78.13%	47.10%
		F1 (%)	38.37%	51.50%	35.85%	59.02%	33.02%	73.86%	48.60%
	MT-CNN	Pre (%)	40.22%	49.63%	70.59%	62.87%	52.78%	72.43%	58.09%
		Rec (%)	36.65%	63.83%	26.36%	60.31%	25.27%	77.81%	48.37%
		F1 (%)	38.36%	55.84%	38.39%	61.56%	34.18%	75.02%	50.56%
	LMT-CNN	Pre (%)	**43.54%**	**50.57%**	**72.92%**	**63.10%**	**53.54%**	**75.91%**	**61.60%**
		Rec (%)	**39.56%**	**64.45%**	**26.47%**	**66.28%**	**28.57%**	**79.32%**	**50.78%**
		F1 (%)	**41.46%**	**56.67%**	**38.84%**	**64.65%**	**37.26%**	**77.58%**	**52.74%**

5. Limitation

In the extensive comparative experiments, the proposed LMT-CNN method shows superior performance on both emotion distribution prediction tasks and emotion classification tasks to the baseline methods. However, our method has some limitations in the following aspects that can be improved in follow-up work.

- The used prior linguistic knowledge is static. Affective words, as a priori knowledge in linguistics, need to build the affective lexicon in advance to associate affective words with emotion labels. However, once the lexicon is established, the affective words have static emotion properties and cannot be updated in real time according to the changes in external emotion semantics. We suppose that dynamic updating of the lexicon is a possible solution.
- The LMT-CNN model does not consider cross-language performance differences. The prior knowledge of affective words is not always present in all languages, especially in

some minor languages. In the future, this deficiency can be addressed through transfer learning, where the rich linguistic knowledge in English is used to boost that of other minor languages.

- Not all state-of-the-art EDL algorithms were directly compared in our experiments. In recent years, many effective EDL models have been proposed, and the algorithms implemented on different datasets are not directly comparable. Re-implementing all state-of-the-art EDL algorithms on their datasets is too time-consuming to illustrate the validity of prior linguistic knowledge. We re-implemented some representative EDL algorithms and performed experimental comparisons on several popular emotional datasets. Although a direct comparison with other algorithms is not possible on different data, classes, and applications, empirical results show that our proposed LMT-CNN method achieves optimal results on all the evaluation metrics.

6. Conclusions

This paper proposes text emotion distribution learning based on the lexicon-enhanced multi-task convolutional neural network (LMT-CNN) model. The LMT-CNN model consists of a semantic information module, emotion knowledge module, and multi-task prediction module. Different from the existing EDL model, LMT-CNN comprehensively considers the linguistic prior knowledge of affective words and text semantic information. A number of comparative experimental results show that the performance of the LMT-CNN model is better than the existing EDL methods in emotion distribution prediction and emotion classification.

In future work, we will consider making more effective use of prior knowledge and try some different emotion modeling methods to predict the emotion distribution.

Author Contributions: Y.D. participated in designing the study, wrote the paper's first draft, performed the computer coding implementation, and performed experiments. X.Z. conceptualized and designed the study, conducted the literature review, and rewrote the paper. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data that support the findings of this study are included in the corresponding referenced articles.

Conflicts of Interest: The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Liu, T.; Du, Y.; Zhou, Q. Text emotion recognition using GRU neural network with attention mechanism and emoticon emotions. In Proceedings of the 2020 2nd International Conference on Robotics, Intelligent Control and Artificial Intelligence, Shanghai, China, 17–19 October 2020; pp. 278–282.
2. Yadollahi, A.; Shahraki, A.G.; Zaiane, O.R. Current state of text sentiment analysis from opinion to emotion mining. *ACM Comput. Surv. (CSUR)* **2017**, *50*, 1–33. [CrossRef]
3. Saraswat, M.; Chakraverty, S.; Kala, A. Analyzing emotion based movie recommender system using fuzzy emotion features. *Int. J. Inf. Technol.* **2020**, *12*, 467–472. [CrossRef]
4. Wang, W.; Zhu, K.; Wang, H.; Wu, Y.-C.J. The impact of sentiment orientations on successful crowdfunding campaigns through text analytics. *IET Softw.* **2017**, *11*, 229–238. [CrossRef]
5. Zhang, L.; Wang, S.; Liu, B. Deep learning for sentiment analysis: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1253. [CrossRef]
6. Geng, X. Label distribution learning. *IEEE Trans. Knowl. Data Eng.* **2016**, *28*, 1734–1748. [CrossRef]
7. Zhou, D.; Zhang, X.; Zhou, Y.; Zhao, Q.; Geng, X. Emotion Distribution Learning from Texts. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–4 November 2016; pp. 638–647.
8. Yang, J.; Li, J.; Li, L.; Wang, X.; Gao, X. A circular-structured representation for visual emotion distribution learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 4237–4246.
9. Xu, Z.; Wang, S. Emotional attention detection and correlation exploration for image emotion distribution learning. *IEEE Trans. Affect. Comput.* **2021**, *1*, 1. [CrossRef]

10. Fei, H.; Zhang, Y.; Ren, Y.; Ji, D. Latent emotion memory for multi-label emotion classification. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 7692–7699.

11. Zhao, Z.; Ma, X. Text emotion distribution learning from small sample: A meta-learning approach. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; pp. 3957–3967.

12. Jia, X.; Zheng, X.; Li, W.; Zhang, C.; Li, Z. Facial emotion distribution learning by exploiting low-rank label correlations locally. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 9841–9850.

13. Pang, J.; Rao, Y.; Xie, H.; Wang, X.; Wang, F.L.; Wong, T.-L.; Li, Q. Fast supervised topic models for short text emotion detection. *IEEE Trans. Cybern.* **2019**, *51*, 815–828. [CrossRef]

14. Teng, Z.; Vo, D.-T.; Zhang, Y. Context-sensitive lexicon features for neural sentiment analysis. In Proceedings of the 2016 conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–5 November 2016; pp. 1629–1638.

15. Zhang, Y.; Fu, J.; She, D.; Zhang, Y.; Wang, S.; Yang, J. Text emotion distribution learning via multi-task convolutional neural network. In Proceedings of the 27th International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 4595–4601.

16. Tong, R.M. An operational system for detecting and tracking opinions in on-line discussion. In Proceedings of the Working Notes of the ACM SIGIR 2001 Workshop on Operational Text Classification, New York, NY, USA; 2001; pp. 1–6.

17. Gao, B.-B.; Xing, C.; Xie, C.-W.; Wu, J.; Geng, X. Deep label distribution learning with label ambiguity. *IEEE Trans. Image Process.* **2017**, *26*, 2825–2838. [CrossRef]

18. Strapparava, C.; Mihalcea, R. Semeval-2007 task 14: Affective text. In Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007), Prague, Czech Republic, 23–24 June 2007; pp. 70–74.

19. Quan, C.; Ren, F. A blog emotion corpus for emotional expression analysis in Chinese. *Comput. Speech Lang.* **2010**, *24*, 726–749. [CrossRef]

20. Alm, C.O.; Sproat, R. Emotional sequencing and development in fairy tales. In Proceedings of the International Conference on Affective Computing and Intelligent Interaction, Beijing, China, 22–24 October 2005; pp. 668–674.

21. Mohammad, S. # Emotional tweets. In Proceedings of the * SEM 2012: The First Joint Conference on Lexical and Computational Semantics–Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012), Montréal, QC, Canada, 7–8 June 2012; pp. 246–255.

22. Scherer, K.R.; Wallbott, H.G. Evidence for universality and cultural variation of differential emotion response patterning. *J. Personal. Soc. Psychol.* **1994**, *66*, 310. [CrossRef]

23. Mohammad, S.; Bravo-Marquez, F.; Salameh, M.; Kiritchenko, S. Semeval-2018 task 1: Affect in tweets. In Proceedings of the 12th International Workshop on Semantic Evaluation, New Orleans, LA, USA, 5–6 June 2018; pp. 1–17.

24. Yao, Y.; Wang, S.; Xu, R.; Liu, B.; Gui, L.; Lu, Q.; Wang, X. The construction of an emotion annotated corpus on microblog text. *J. Chin. Inf. Process.* **2014**, *28*, 83–91.

25. Zhao, S.; Ding, G.; Huang, Q.; Chua, T.-S.; Schuller, B.W.; Keutzer, K. Affective image content analysis: A comprehensive survey. In Proceedings of the 27th International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 5534–5541.

26. Zhao, S.; Ding, G.; Han, J.; Gao, Y. Personality-aware personalized emotion recognition from physiological signals. In Proceedings of the 27th International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 1660–1667.

27. Ekman, P. Basic emotions. *Handb. Cogn. Emot.* **1999**, *98*, 16.

28. Mikels, J.A.; Fredrickson, B.L.; Larkin, G.R.; Lindberg, C.M.; Maglio, S.J.; Reuter-Lorenz, P.A. Emotional category data on images from the International Affective Picture System. *Behav. Res. Methods* **2005**, *37*, 626–630. [CrossRef]

29. Schlosberg, H. Three dimensions of emotion. *Psychol. Rev.* **1954**, *61*, 81. [CrossRef]

30. Perikos, I.; Hatzilygeroudis, I. Aspect based sentiment analysis in social media with classifier ensembles. In Proceedings of the 2017 IEEE/ACIS 16th International Conference on Computer and Information Science, Wuhan, China, 24–26 May 2017; pp. 273–278.

31. Vrysis, L.; Vryzas, N.; Kotsakis, R.; Saridou, T.; Matsiola, M.; Veglis, A.; Arcila-Calderón, C.; Dimoulas, C. A Web Interface for Analyzing Hate Speech. *Future Internet* **2021**, *13*, 80. [CrossRef]

32. Machova, K.; Mach, M.; Vasilko, M. Comparison of Machine Learning and Sentiment Analysis in Detection of Suspicious Online Reviewers on Different Type of Data. *Sensors* **2021**, *22*, 155. [CrossRef] [PubMed]

33. Khan, I.U.; Khan, A.; Khan, W.; Su'ud, M.M.; Alam, M.M.; Subhan, F.; Asghar, M.Z. A Review of Urdu Sentiment Analysis with Multilingual Perspective: A Case of Urdu and Roman Urdu Language. *Computers* **2021**, *11*, 3. [CrossRef]

34. Catelli, R.; Pelosi, S.; Esposito, M. Lexicon-based vs. Bert-based sentiment analysis: A comparative study in Italian. *Electronics* **2022**, *11*, 374. [CrossRef]

35. Franzoni, V.; Milani, A.; Biondi, G. SEMO: A semantic model for emotion recognition in web objects. In Proceedings of the International Conference on Web Intelligence, Leipzig, Germany, 23–26 August 2017; pp. 953–958.

36. Dashtipour, K.; Gogate, M.; Cambria, E.; Hussain, A. A novel context-aware multimodal framework for Persian sentiment analysis. *Neurocomputing* **2021**, *457*, 377–388. [CrossRef]

37. Rahman, W.; Hasan, M.K.; Lee, S.; Zadeh, A.; Mao, C.; Morency, L.-P.; Hoque, E. Integrating multimodal information in large pretrained transformers. In Proceedings of the Conference of Association for Computational Linguistics. Meeting, Seattle, WA, USA, 5–10 July 2020; p. 2359.

38. Kydros, D.; Argyropoulou, M.; Vrana, V. A content and sentiment analysis of Greek tweets during the pandemic. *Sustainability* **2021**, *13*, 6150. [CrossRef]

39. Chen, T.; Yin, X.; Peng, L.; Rong, J.; Yang, J.; Cong, G. Monitoring and recognizing enterprise public opinion from high-risk users based on user portrait and random forest algorithm. *Axioms* **2021**, *10*, 106. [CrossRef]

40. Zou, H.; Xiang, K. Sentiment classification method based on blending of emoticons and short texts. *Entropy* **2022**, *24*, 398. [CrossRef] [PubMed]

41. Socher, R.; Perelygin, A.; Wu, J.; Chuang, J.; Manning, C.D.; Ng, A.Y.; Potts, C. Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, WA, USA, 18–21 October 2013; pp. 1631–1642.

42. Zheng, X.; Jia, X.; Li, W. Label distribution learning by exploiting sample correlations locally. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018.

43. Batbaatar, E.; Li, M.; Ryu, K.H. Semantic-emotion neural network for emotion recognition from text. *IEEE Access* **2019**, *7*, 111866–111878. [CrossRef]

44. Qin, X.; Chen, Y.; Rao, Y.; Xie, H.; Wong, M.L.; Wang, F.L. A constrained optimization approach for cross-domain emotion distribution learning. *Knowl.-Based Syst.* **2021**, *227*, 107160. [CrossRef]

45. Wang, Y.; Pal, A. Detecting emotions in social media: A constrained optimization approach. In Proceedings of the Twenty-fourth International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015.

46. Abdi, A.; Shamsuddin, S.M.; Hasan, S.; Piran, J. Deep learning-based sentiment classification of evaluative text based on Multi-feature fusion. *Inf. Process. Manag.* **2019**, *56*, 1245–1259. [CrossRef]

47. Ke, P.; Ji, H.; Liu, S.; Zhu, X.; Huang, M. SentiLARE: Sentiment-aware language representation learning with linguistic knowledge. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, Online, 16–20 November 2020; pp. 6975–6988.

48. Robbins, H.; Monro, S. A stochastic approximation method. *Ann. Math. Stat.* **1951**, *22*, 400–407. [CrossRef]

49. Lexicon-Enhanced Multi-Task Convolutional Neural Network for Emotion Distribution Learning. Available online: https://github.com/yc-Dong/Lexicon-enhanced-multi-task-CNN (accessed on 19 December 2021).

50. Poria, S.; Gelbukh, A.; Cambria, E.; Hussain, A.; Huang, G.-B. EmoSenticSpace: A novel framework for affective common-sense reasoning. *Knowl. -Based Syst.* **2014**, *69*, 108–123. [CrossRef]

51. Mohammad, S.; Turney, P. *Nrc Emotion Lexicon*; National Research Council: Ottawa, ON, Canada, 2020; Volume 2013, p. 2.

52. Xu, L.; Lin, H.; Pan, Y.; Ren, H.; Chen, J. Construction of emotional vocabulary ontology. *J. China Soc. Sci. Tech. Inf.* **2008**, *27*, 180–185.

53. Kim, Y. Convolutional neural networks for sentence classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 25–29 October 2014; pp. 1746–1751.

54. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. *Adv. Neural Inf. Process. Syst.* **2013**, *26*, 3111–3119.

55. Li, S.; Zhao, Z.; Hu, R.; Li, W.; Liu, T.; Du, X. Analogical reasoning on Chinese morphological and semantic relations. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Melbourne, Australia, 15–20 July 2018; pp. 138–143.

56. Molchanov, P.; Tyree, S.; Karras, T.; Aila, T.; Kautz, J. Pruning convolutional neural networks for resource efficient inference. In Proceedings of the 5th International Conference on Learning Representations, ICLR 2017-Conference Track Proceedings, Toulon, France, 24–26 April 2017.