# ARTICLE

https://doi.org/10.1057/s41599-023-01816-6

OPEN

# Check for updates

# Emotion classification for short texts: an improved multi-label method

Xuan Liu<sup>® 1</sup>, Tianyi Shi<sup>2</sup>, Guohui Zhou<sup>1</sup>, Mingzhe Liu<sup>3</sup>, Zhengtong Yin<sup>4</sup>, Lirong Yin<sup>5</sup> & Wenfeng Zheng<sup>2⊠</sup>

The process of computationally identifying and categorizing opinions expressed in a piece of text is of great importance to support better understanding and services to online users in the digital environment. However, accurate and fast multi-label automatic classification is still insufficient. By considering not only individual in-sentence features but also the features in the adjacent sentences and the full text of the tweet, this study adjusted the Multi-label K-Nearest Neighbors (MLkNN) classifier to allow iterative corrections of the multi-label emotion classification. It applies the new method to improve both the accuracy and speed of emotion classification for short texts on Twitter. By carrying out three groups of experiments on the Twitter corpus, this study compares the performance of the base classifier of MLkNN, the sample-based MLkNN (S-MLkNN), and the label-based MLkNN (L-MLkNN). The results show that the improved MLkNN algorithm can effectively improve the accuracy of emotion classification of short texts, especially when the value of K in the MLkNN base classifier is 8, and the value of  $\alpha$  is 0.7, and the improved L-MLkNN algorithm outperforms the other methods in the overall performance and the recall rate reaches 0.8019. This study attempts to obtain an efficient classifier with smaller training samples and lower training costs for sentiment analysis. It is suggested that future studies should pay more attention to balancing the efficiency of the model with smaller training sample sizes and the completeness of the model to cover various scenarios.

<sup>&</sup>lt;sup>1</sup>School of Public Affairs and Administration, University of Electronic Science and Technology of China, 610054 Chengdu, China. <sup>2</sup>School of Automation, University of Electronic Science and Technology of China, 610054 Chengdu, China. <sup>3</sup>School of Data Science and Artificial Intelligence, Wenzhou University of Technology, 325000 Wenzhou, China. <sup>4</sup>College of Resource and Environment Engineering, Guizhou University, 550025 Guiyang, China. <sup>5</sup>Department of Geography and Anthropology, Louisiana State University, Baton Rouge, LA 70803, USA. <sup>53</sup>email: winfirms@uestc.edu.cn

# Introduction

ast-increasing electronic documents in the digital environment offer a new source to support better understanding and services to online users. More attention has been paid to extracting users' opinions towards various events from the content of texts. The process of computationally identifying and categorizing opinions expressed in a piece of text has been highlighted as the first step of data mining. Scholars tried to extract the positive, negative, or neutral attitude toward a particular topic or product from the text (Feng et al., 2021) and recently started to further label the texts with multi-dimensional emotional tendencies, such as joy, fear, rage, etc. (Hu and Flaxman, 2018; Tasmin, 2018). Various methods have been applied to classify the emotions of the texts, but the methods based on machine learning have attracted the most attention (Chen and Zhang, 2018). Previous methods based on the dictionaries of emotions allow segmentation and classification of words for analysis of complicated emotions, preparing a dictionary of emotions is labor-intensive and time-consuming and could hardly catch up with the fast emergence of new words (Ai et al., 2018). Machine-learning algorithms, on the contrary, allow auto recognition of emotional words in texts so as to achieve classification more quickly. However, the sequential process that the machine-learning algorithms follow would inevitably result in the inability to label multiple emotions and the possibility of the heavy impact of previous steps on the following steps (Ullah et al., 2022). Problems such as the decline of classifier performance with emotion refinement, the lack of the relationship between sentences and the whole text, and the recognition of complex human emotions are also stimulating scholars to keep adjusting the algorithms to enhance their performance.

One of the key directions to improve the machine-learning algorithms for emotion classification is computational multi-label classification. Computational multi-label classification is regarded as a good solution. Multi-label classification means that an instance could be classified into multiple categories at the same time; that is, it could be marked by multiple labels. In practical applications, the semantics of real objects or real texts are often not unique, which leads to the need for multi-label learning. Mainly by proposing new emotional dictionaries, some pioneering research has made remarkable attempts in the field of multi-label classification (Yang et al., 2014; Liu and Chen, 2015). However, it becomes very difficult for algorithms to classify emotions with multiple labels. Compared with SVM and Bayesian algorithms, K-Nearest Neighbors (KNN) algorithms perform best as a multi-label classifier and are easier to construct (Keshtkar and Inkpen, 2012). The problem remains that the iterative corrections are not able to be achieved for emotion classification even for kNN algorithms.

In response to the above knowledge gaps, this study adjusts the Multi-label K-Nearest Neighbors (MLkNN) classifier and considers not only individual in-sentence features but also the features in the adjacent sentences and the full text of the tweet. Furthermore, this study further considers the interaction between labels and iteratively updates the overall classification results. Such adjustments allow iterative corrections of the multi-label emotion classification and could improve the accuracy of emotion classification for short texts. Tweets are chosen in this study as a representational source of short texts. Among all text classifications, short text classification is a special subdomain with increasing importance. Since people are now more frequently using short sentences to express opinions or share ideas with others, short text classification becomes essential in author recognition, spam filtering, sentiment analysis, Twitter personalization, customer review, and other applications related to social networks (Liang et al., 2020). Therefore, there is an expanding need for sentiment analysis of short texts on the internet.

The rest of the study is organized as follows: studies about existing emotion classification methods in the literature are summarized in section "Related work". The adjustments to the MLkNN emotion analysis method are employed in section "An improved MLkNN for emotion classification of short texts", followed by the experiments and results in section "Experimental study". Lastly, the discussions and conclusions are presented in sections "Discussion" and "Conclusions", respectively.

# **Related work**

Human emotion has been a research hotspot for scholars since ancient times. In the era of information, emotional signals and sentiment tendencies also attract more attention when scholars extract textual features from the content of online texts to support better understandings and services to users. The specific task of sentiment classification is to identify the subjective views expressed in the specified text and judge the emotional tendencies of the text (Rajabi et al., 2020; Li et al., 2016; Fei et al., 2020). Together with the accumulation of the understanding of texts, two types of emotion classifications have been highlighted. One is the classification of emotions according to the emotional polarities, which means the positive, negative, or neutral attitudes. The other one is the classification of emotions according to emotional tendencies, which generally follows the emotion wheel proposed by American psychologist Plutchik (Hu and Flaxman, 2018; Tasmin, 2018). The introduction of emotional tendencies increases of the emotion classifications and leads to a review of the emotion classification methods.

Among the emotion classification methods, the most used and better-performing methods are mostly based on dictionaries of emotions and based on machine learning (Chen and Zhang, 2018). Emotion classification by a dictionary of emotions is a classical method with both theoretical and practical achievements (Ai et al., 2018). The main implementation processes of emotion classification include segmenting words in the text to be classified and carrying out keyword matching and other operations on these words to realize emotion classification. Ma et al. (2005) first applied the dictionary-based method to the instant messaging system. On this basis, Aman and Szpakowicz (2007) proposed a classification method by adding the emotion intensity knowledge base to the original dictionary and achieved an accuracy of more than 66% in the emotion classification task of the blog corpus. Paltoglou and Thelwall (2012) used the dictionary of emotions method to calculate the negative words, capital letters, emotional polarity, and their strength changes in the linguistic field. The accuracy rate of this method could reach 86.5% when applied to short texts on platforms such as Twitter and MySpace. Taboada et al. (2011) further expanded the dictionary of emotional features and topic-related features of the text. The accuracy of the improved method in the experiment of the Twitter corpus could reach 85.6%.

Due to the long training time of many networks, some researchers have defined several dictionaries related to emotional words, such as attitude dictionary, negative dictionary, degree dictionary and connective dictionary. In addition, there are more complex methods in recent years, such as the emotion classification method based on rules (Yan et al., 2018). These works include the classification method based on mutual information (Liu et al., 2021), the emotion classification method based on physiological signals (Shu et al., 2018), as well as the upgrading of neural networks (Tang et al., 2021). The classification method based on a dictionary of emotions could reflect the unstructured features of the text and have a high utilization rate of emotional words, but its problems are also obvious: the scarcity of corpus

resources, the low update frequency of emotional words, and the inability to capture new words or deformed words. Ideal classification requires higher rates of coverage and labeling accuracy of emotional words in the dictionary. Moreover, dictionaries are highly dependent on the domain, time, language and other conditions, thus difficult to expand.

In recent years, the rapid development of machine-learning methods offers a new way for emotion text classification. Two types of schemes have been applied: supervised and semisupervised. The common features used for emotion classification in the existing supervised learning schemes mainly include wordlevel, sentence-level and chapter-level features (Dogan and Uysal, 2020). Keshtkar and Inkpen (2012) adopted multi-level analysis ideas to analyze the mood of bloggers and achieved stratified emotion analysis for more than 100 mood labels. Semi-supervised learning schemes differ from supervised learning in that they require a large number of labeled samples. Semi-supervised learning could utilize a large number of unlabeled samples, which could improve the classifier performance and reduce the dependence on sample sets. Presently, the existing semi-supervised learning methods in the field of emotion classification include the semi-supervised emotion classification schemes based on multinomial Bayes, discrete binary semi-supervised learning, Emoji space model, and dual view label propagation (Sintsova et al., 2014). The main advantages of this method are that it does not depend on a large number of labeled samples and could easily obtain a large number of new labeled data as training samples through learning. It performs well when the scarcity of labeled datasets occurs. However, the disadvantages of this method are also obvious. It is very sensitive to the results of the first round of classification: the samples that could not be correctly classified in the first classification process will greatly affect the accuracy of the second classification.

New neural networks have also been applied as feature extractors. Liao et al. (2021) proposed a novel two-stage finegrained text-level sentiment analysis model based on syntactic rule matching and deep semantics. Combining the multi-head attention mechanism in Transformer, Lou et al. (2020) proposed a fusion model of convolution neural network and hierarchical attention coding network to avoid the sequential processing of Recurrent Neural Networks (RNN), which were wildly used as the feature extractor for fine-grained sentiment analysis. The selfattention-based Bidirectional Long Short-Term Memory (BiLSTM) model with aspect item information for fine-grained sentiment classification of short texts introduced by Xie et al. (2019) allowed effective use of contextual information and semantic features. A recent study of a bidirectional convolutional RNN adopted bidirectional feature extraction to group features and enhanced the important features in each group while weakening the less important features to improve the classification accuracy (Onan, 2022). Jiang et al. (2022) mixed Bidirectional Encoder Representation from Transformers (BERT), BiLSTM and Text Convolutional Neural Network (TextCNN) into a new model, achieving not only the capture of local correlations in contexts, but also high accuracy and stability. However, machinelearning methods still face the following defects. One is that they often rely excessively on the manually labeled corpus, and could not achieve good results when the sample set size is small. Besides, unsupervised learning in machine learning is still scarce in the field of sentiment analysis.

Recently, scholars started to realize that it is often unable to accurately restore and analyze the individual's real emotions without considering of multiple emotions contained in the text (Siriwardhana et al., 2020; Sadr et al., 2019; Liang et al., 2019). Multi-label learning originated from the investigation of text classification problems, where each document may belong to several predefined topics simultaneously. In multi-label learning, the training set is composed of instances each associated with a set of labels, and the task is to predict the label sets of unseen instances by analyzing training instances with known label sets (Zhang and Zhou, 2007). Yang et al. proposed a small dictionary that considers not only text, but also graphic emoticons and punctuation marks, and classify multi-label sentiment of Weibo corpus. This classification has achieved a relatively high accuracy rate and has played an active role in the analysis of public opinion on the Malavsia Airlines crash. Liu and Chen (2015) used the combination of three emotional dictionaries to extract the emotional features and the original segmentation word features in the microblog corpus, and completed a multi-label-based emotion classification method, among which the best experimental results had an average accuracy of 65.5%. In response to the imbalance of emotion category distribution in the corpus, Li et al. (2016) adopted a multi-label maximum entropy model to analyze the relationship between words and emotions. The problem remains as iterative corrections are not able to be achieved for the emotion classification. These methods thus could hardly catch up with the fast changes of emotional words in the real world.

An improved MLkNN for emotion classification of short texts Introduction of multi-label classification and the workflow. In machine learning, multi-label classification or multi-output classification is a variant of the classification problem where multiple non-exclusive labels may be assigned to each instance. In the multi-label problem, the labels are non-exclusive and there is no constraint on how many of the classes the instance could be assigned to. The definition of multi-label problems could be specifically explained by the following mathematical notation:

Let  $X = \{x_1, x_2, \dots, x_n\}$  represent the example space,  $L = \{L_1, L_2, \dots, L_q\}$  represent the set of all labels and  $Y = \{y_1, y_2, \dots, y_n\}$  represent the label space. To complete the multi-label classification, the first task is to obtain the function  $f: X \rightarrow Y$  by learning from the training set  $\{(x_i, y_i) | 1 \le i \le m\}$ , where  $x_i \in X$  is an example,  $y_i \in Y$  is the category label to which the example  $x_i$  belongs, and  $y_i$  is a subset of the label set L. In practical applications, real objects or real text are often not unique in the semantic language, thus leading to a multi-label learning framework. In the framework of multi-labels, each object is formulated by an instance with multiple category labels, and the goal of learning is to assign all appropriate labels to that instance.

To solve the multi-label classification problem, the mainstream solutions include three different ways: problem transformation methods, adapted algorithms, and integration methods. Adapted algorithms have attracted the most attention from scholars as more choices are available: *K*-nearest neighbor (KNN) (Zhang and Zhou, 2007), multi-valued and multi-labeled decision tree (Chou and Hsu, 2005), kernel methods for vector output and neural networks such as BP-MLL (Zhang and Zhou, 2006) could all be applied for a better solution of the multi-label classification problem.

This paper focuses on modifying the ML-kNN algorithm through the following steps: first, the short text is divided according to the sentence, the emotion transfer relationship characteristics between adjacent sentences and the emotion transfer relationship characteristics between sentences and in the full text of the tweet are gained through the training set. MLkNN multi-label classifier is thus applied as the base classifier for the emotion classification to get the initial classification results of sentences and calculate the results of the overall emotion classification results of tweets for the test set, and then modify the overall emotion classification results of tweet specifically by the rate of emotion transfer relationship between sentences and emotion transfer in the full text of the tweet. Average Precision



Fig. 1 Workflow of the study.

(AVP) is thus evaluated to decide whether the process goes to further adjustment based on label correlation or back to the steps of adjustments based on emotion transfer. The workflow of the study is shown in Fig. 1.

This method would be more applicable for tweets, which is generally short and the expression is usually colloquial. The features in sentences could hardly support the judgment of the sentence emotional category. Deviation occurs for sentences that only contain oral vocabulary.

Actually, the data supporting the results of this study were downloaded from www.kaggle.com, which provided sentiment140 data earlier than 2009 and was not affected by the increase in the length of tweets in 2017. In this corpus, tweets are mostly in the same form as the following example. There are three short sentences in this example, all of which express "joy". However, the first sentence is often used to express strong emotions on the Internet, and in many cases, it is used to express negative emotions such as anger or disgust. If only the features of the words in the sentence are considered, the emotion of the first sentence is likely to be classified into the category of "disgust" or "anger". However, if the emotional transfer caused by the context of the sentence is considered at the same time, and the overall mood of Twitter is considered, the mood category of the first sentence could be modified in the case that both the adjacent sentences are joy mood.

<sentence ID= "1">WTF! </sentence>

<sentence ID= "2">This song is groovy!! </sentence>

<sentence ID= "3">I love the song soooo much! </sentence>

#### </tweet>

The suitable way for the emotional classification of a tweet is to consider the emotional transfer relationship of adjacent sentences and the full text of the tweet to reduce the error caused by text ambiguity and to improve the overall classification accuracy. **Base classifier based on MLkNN**. MLkNN algorithm is derived from the traditional *K*-nearest neighbor (KNN) algorithm. This method finds the label information contained in the *K*-nearest neighbor instances of the target, following statistical methods, infers the label set of the target by maximizing the posterior probability (Zhang and Zhou, 2007). When using the MLkNN algorithm for text sentiment classification, the specific calculation method is as follows. Let L be the set of emotional labels, for an emotional label  $l \in L$ , the event is defined as:

 $Y_l^0$ : Text s does not contain the emotion label *l*.

 $Y_l^{i_1}$ : Text s contains the emotion label *l*.

 $H_l^t$ ,  $(t \in \{0, 1, \dots, k\})$ : There are exactly t texts containing label l among the K-nearest neighbors N(s) of texts.

According to the maximum posterior criterion, the category vector  $\vec{y}_{s}(l)$  is defined as follows:

$$\vec{y}_s(l) = \arg\max_{e \in \{0,1\}} P\left(Y_l^e \big| H_l^{\vec{C}_s(l)}\right) \tag{1}$$

In Formula 1,  $\vec{C}_s(l) = \sum_{a \in N(s)} \overrightarrow{y}_a(l)$  is the number of nearest neighbors containing label l in the *K*-nearest neighbors of text s. According to the Bayesian theory, Formula 1 could be further rewritten as:

$$\vec{y}_{s}(l) = \arg \max_{e \in \{0,1\}} \frac{P(Y_{l}^{e}) P(H_{l}^{\vec{C}_{s}(l)} | Y_{l}^{e})}{P(H_{l}^{\vec{C}_{s}(l)})} = \arg \max_{e \in \{0,1\}} P(Y_{l}^{e}) P(H_{l}^{\vec{C}_{s}(l)} | Y_{l}^{e})$$
(2)

In Formula 2, the prior probabilities  $P(Y_l^e)$  and the probability  $P(H_l^t|Y_l^e)$  could be calculated from the training set sepcifically, as shown in Formulas 3, 4 and 5:

$$P(Y_l^1) = (s + \vec{C}_s(l))/(s + 2 + m); P(Y_l^0) = 1 - P(Y_l^e)$$
(3)

$$P(H_l^t | Y_l^1) = (s + c[t]) / \left(s + (k+1) + \sum_{p=0}^k c[p]\right)$$
(4)

$$P(H_l^t | Y_l^0) = (s + c'[t]) / \left(s + (k+1) + \sum_{p=0}^k c'[p]\right)$$
(5)

The algorithm presented in this section is an improvement to the MLkNN classifier. For the initial emotion classification of sentences, the improved algorithm uses the in-sentence features as the initial features and combines the MLkNN algorithm to construct the sentence base classifier. The initial result of the emotional classification of the sentence in a tweet is obtained with Formula 6. Both the prior and conditional probabilities of each emotion could be obtained through the training set, as shown in Formulas 3, 4 and 5.

$$\vec{y}_a(l) = \arg \max_{e \in \{0,1\}} P(Y_l^e) P\left(H_l^{\vec{\zeta}_S(l)} | Y_l^e\right)$$
(6)

Adjustment with the probability of emotional transfer between sentences. After the initial classification results of sentences obtained by MLkNN classifier, the classification accuracy of sentences is not yet ideal because only the in-sentence features are considered. This study believes that the emotion category of sentences could be further modified by using the emotion transfer probability of adjacent sentences. We mark the first sentence of sentence s as  $S_P$  and the last sentence of sentence s as  $S_N$ . Possible events are defined as follows:

- $P_{\varepsilon}^{1}$ : In the previous sentence of the sentence s,  $S_{P}$ , the emotion is  $\varepsilon$ .
- $P_{\varepsilon}^{0}$ : In the previous sentence of the sentence s,  $S_{P}$ , the emotion is not  $\varepsilon$ .
- $N_{\xi}^{1}$ : In the next sentence of a sentence s,  $S_{N}$ , the emotion is  $\varepsilon$ .
- $N_{\varepsilon}^{\xi_0}$ : In the next sentence of a sentence s,  $S_N$ , the emotion is not  $\varepsilon$ .

Assuming that the emotional transfers between any two adjacent sentences in the tweet are independent of each other,  $\vec{y}_{s}(l)$  is defined in Formula 7:

$$\vec{y}_{s}(l) = \arg \max_{e \in \{0,1\}} P\left(Y_{l}^{e} | H_{l}^{\vec{C}_{s}(l)}, P_{\varepsilon}^{\vec{\gamma}(\varepsilon)} \cdots, N_{\varepsilon}^{\vec{\gamma}(\varepsilon)} \cdots\right)$$

$$= \arg \max_{e \in \{0,1\}} P(Y_{l}^{e}) P\left(H_{l}^{\vec{C}_{s}(l)} | Y_{l}^{e}\right) \cdot \prod_{\varepsilon \in L} P\left(P_{\varepsilon}^{\vec{\gamma}(\varepsilon)} | Y_{l}^{e}\right) \cdot P\left(N_{\varepsilon}^{\vec{\gamma}(\varepsilon)} | Y_{l}^{e}\right)$$
(7)

The calculation with Formula 7 requires the emotion transfer probability between adjacent sentences. To achieve that, the transition probability from emotion x to emotion l could be calculated from the sample set, and the emotion transfer probability could be obtained according to Formula 8 in which the emotional transition probabilities with its previous and next sentences are calculated respectively.

$$P(P_{\varepsilon}^{1}|Y_{l}^{1}) = p(\varepsilon \to l) = \frac{count(\vec{y}_{sp}(\varepsilon)=1, \vec{y}_{s}(l)=1)}{count(\vec{y}_{s}(l)=1)}$$

$$P(N_{\varepsilon}^{1}|Y_{l}^{1}) = p(\varepsilon \to l) = \frac{count(\vec{y}_{sn}(\varepsilon)=1, \vec{y}_{s}(l)=1)}{count(\vec{y}_{s}(l)=1)}$$
(8)

Adjustment with the probability of emotional transfer in a full text of the tweet. Similar to the calculation method of the emotion transfer probability between adjacent sentences, the emotion transition probability can be calculated for the overall emotion of the tweet and the emotion of the sentence. Assuming that w is the tweet where the sentence s is located, the possible event is defined as follows:

- $W^1_{\varepsilon}$ : The emotion of the tweet w where the sentence s contains  $\varepsilon$ .
- $W^0_\varepsilon$  : The emotion of the tweet w where the sentence s does not contain  $\varepsilon.$

Assuming that the emotional transformations between the full text of the tweet and each sentence are independent events,  $\vec{y}_s(l)$  could be defined with Formula 9.

$$\vec{y}_{s}(l) = \arg \max_{e \in \{0,1\}} P\left(Y_{l}^{e} \middle| H_{l}^{\vec{C}_{s}(l)}, W_{\varepsilon}^{\vec{y}(e)} \cdots\right)$$

$$= \arg \max_{e \in \{0,1\}} P\left(Y_{l}^{e}\right) P\left(H_{l}^{\vec{C}_{s}(l)} \middle| Y_{l}^{e}\right) \cdot \prod_{\varepsilon \in L} P\left(W_{\varepsilon}^{\vec{y}(\varepsilon)} \middle| Y_{l}^{e}\right)$$
(9)

The emotion transfer probability from  $\varepsilon$  to l could be calculated by Formula 10:

$$p(\varepsilon \to l) = \frac{\operatorname{count}\left[\vec{y}_w(\varepsilon) = 1, \, \vec{y}_s(l) = 1\right]}{\operatorname{count}\left[\vec{y}_s(l) = 1\right]} \tag{10}$$

**Adjustment to MLkNN**. The classification model based on MLkNN regards that there is no relationship between multiple labels of the same text, and does not consider the influence of the relationship between its labels for each instance. To improve that, the label correlation of the instance itself is taken as a related factor of multi-label emotion classification to further enhance the accuracy. First, we investigate the correlation of multiple labels of each instance in the training set. The existing solution strategies to investigate the relevance of labels could be roughly divided into three categories according to the complexity of calculation (Zhang and Zhang, 2010; Yang et al., 2019):

(1) The first-order strategy examines each label in turn and decomposes the multi-label learning problem into an independent binary classification problem. This method is easy to implement, but difficult to generalize.

- (2) The second-order strategy examines each case of label pairwise combination. This method takes into account the correlation between labels but does not include all inclusion cases of labels.
- (3) High-order strategy, which investigates the high-order correlation between labels, is more comprehensive than the above two. However, it also brings high computational complexity, which is difficult to be applied to large-scale learning problems.

Taking into consideration that there is often correlation between multiple sentiment tags of a text, this study adopts the second-order strategy. For all labels set  $L=\{L_1, L_2, \dots, L_q\}$ , the total number of possible combinations is q(q-1)/2.  $L_{new}$  is defined as  $|L_{new}| = q(q+1)/2$ . For any instance in the training set, the corresponding 0 / 1 labels for  $L_1$  to  $L_q$  remain the same as before expansion. The label sets from  $L_{q+1}$  to  $L_{q(q+1)/2}$  need further adjustment: we label 0 for those with two labels  $L_i$  and  $L_j(1 \le i \le q, 1 \le j \le q, i \ne j)$ , otherwise, 0.

Word co-occurrence patterns proposed by Baeza-Yates and Ribeiro-Neto (1999) are applied as a measure of emotion correlation between labels for automatic local analysis. The cooccurrence matrix M composed by the emotion labels and instances is shown in Table 1.

In the table,  $x_i(1 \le i \le n)$  represents the i-th sample in the training set and  $\omega_j(1 \le j \le q)$  represents all data from row j in the matrix M. By normalizing them, the common co-occurrence frequency of the labels  $L_u$  and  $L_v$  is shown in Formulas 11 and 12:

$$C_{\omega u,\omega v} = \sum M_{ui} \times M_{vi} \tag{11}$$

$$S_{\omega u,\omega v} = \frac{C_{\omega u,\omega v}}{C_{\omega u,\omega u} + C_{\omega v,\omega v} - C_{\omega u,\omega v}}$$
(12)

Here,  $S_{\omega u,\omega v}$  indicates the frequency of  $\omega u$  and  $\omega v$  cooccurrence and is shown as a symmetry matrix  $Q1_{ij}$  with all the diagonal elements as 1. *YY1(i)* and *YY0(i)* are obtained with Formulas 13 and 14, in which PHY1 and PHY2 are obtained with Formulas 3 and 4.

$$YY1(i) = \sum Q1_{ii} \times PHY1(ij,d)$$
(13)

$$YY0(i) = \sum Q0_{ii} \times PHY0(ij,d) \tag{14}$$

Assuming that a parameter  $\alpha$  satisfies  $0 \le \alpha \le 1$ , the previously obtained YY1(i) and YY0(i) are combined with Y1(i) and Y0(i) calculated in MLkNN. y1(i) and y0(i) are obtained with Formulas 15 and 16 to judge whether the test instance contains the label  $L_i$ .

$$y1(i) = \alpha \times Y1(i) + (1 - \alpha) \times YY1(i)$$
(15)

$$y0(i) = \alpha \times Y0(i) + (1 - \alpha) \times YY0(i)$$
(16)

The sentence-level emotion classification results of the tweet obtained in section "Adjustment with the probability of emotional transfer in a full text of the tweet" are integrated and calculated according to Formula 17 to obtain the full text of the tweet-level emotion classification results, which are then used as

Table 1 Co-occurrence matrix M.									
		<i>x</i> 1	x <sub>2</sub>	<b>X</b> 3	<b>x</b> 4		x <sub>n</sub>		
$\omega_1$	$L_1$	0	1	0	1		1		
ω2	L <sub>2</sub>	1	0	1	0		1		
$\omega_3$	L <sub>3</sub>	1	0	1	1		0		
:	:	:	:	:	:	:	:		
$\omega_q$	L <sub>q</sub>	1	1	0	0		1		

the basic classification results in the improved algorithm for further improvement.

$$\vec{y}_{w}(l) = \frac{1}{p} \sum_{i=0}^{p} \vec{y}_{s}(l)$$
(17)

Here, p is the total number of sentences in each tweet.

# **Experimental study**

**Experimental settings.** The Sentiment140 Twitter corpus<sup>1</sup> contains 1,600,000 tweets extracted using the Twitter API (Go et al., 2009). In this study, 8000 Twitter texts were randomly selected from Sentiment140 for annotation. After filtering the meaningless text, 6500 Twitter texts were finally retained, containing a total of 11,338 sentences, which cover a wide range of content and have the common text characteristics of short Internet texts. Each complete tweet instance contains up to two emotion labels, and each sentence in the instance contains up to one emotion label. According to the experimental needs, the dataset is divided into two parts: training set and test set, with a corpus ratio of 7:3. The training set contains 4500 tweet data and 7779 sentences, and the test set contains 2000 tweet data and 3559 sentences.

The multi-label classification evaluation indicators are mainly divided into two categories: sample-based and label-based indicators. The sample-based indicators mainly consider the evaluation results of each sample, and then take the average of multiple samples. The label-based indicators mainly consider the performance of a single label on all samples, and then take the average of multiple labels. The experiments in this section mainly evaluate the performance of multi-label classification with the sample-based indicators, which include Subset Accuracy (SA), Hamming Loss (HL), One-Error (OE), Ranking Loss (RL), Average Precision (AVP), Accuracy (AC), Precision (PR), Recall (RE) and F-score. Specifically, SA is the parameter that measures the accuracy rate. HL measures the proportion of misclassified labels. OE refers to the proportion of samples predicting the most relevant labels that are not present in the real labels. RL indicates the situation when the correlations of uncorrelated labels are scored higher than related labels. AVP means the proportion of the predictions when relevant labels rank higher than a chosen label. Accuracy, Precision, Recall and F-score are the extension of accuracy, precision, recall and F-value in the single-label classification task. The calculation methods are introduced respectively:

$$SA = \frac{1}{p} \sum_{i=1}^{p} 1\{h(x^i) = y^i\}$$
(18)

*p* indicates the sample size of the test set  $1{\pi}$  returns 1 when  $\pi$  is true and 0 otherwise.

$$HL = \frac{1}{p} \sum_{i=1}^{p} \frac{1}{q} \left| h\left(x^{i}\right) \Delta y^{i} \right|$$
(19)

*q* represents the total number of all labels,  $\Delta$  means exclusive OR,  $y^i$  represents the set of actual labels of sample *i*, and  $h(x^i)$  represents the set of predicted labels of sample *i*.

$$OE = \frac{1}{p} \sum_{i=1}^{p} 1\left\{ \left[ \arg \max_{y_i \in Y} f\left(x^i, y_j\right) \right] \notin y^i \right\}$$
(20)

$$RL = \frac{1}{p} \sum_{i=1}^{p} \frac{1}{|y^{i}| |\overline{y^{i}}|} \left\{ \begin{cases} \left(y_{j1}, y_{j2}\right) \left| f\left(x^{i}, y_{j1}\right) \le f\left(x^{i}, y_{j2}\right), \\ \left(y_{j1}, y_{j2}\right) \in \left(y^{i} \times \overline{y^{i}}\right) \end{cases} \right\}$$
(21)

$$AVP = \frac{1}{p} \sum_{i=1}^{p} \frac{1}{|y^{i}|} \sum_{y_{j_{1}} \in y^{i}} \frac{\left| \left\{ y_{j_{2}} \middle| rank_{f} \left( x^{i}, y_{j_{2}} \right) \le rank_{f} \left( x^{i}, y_{j_{1}} \right), y_{j_{2}} \in y^{i} \right\} \right|}{rank_{f} \left( x^{i}, y_{j_{1}} \right)}$$
(22)

$$AC = \frac{1}{p} \sum_{i=1}^{p} \frac{|h(x^{i}) \cap y^{i}|}{|h(x^{i}) \cup y^{i}|}$$
(23)

$$PR = \frac{1}{p} \sum_{i=1}^{p} \frac{|h(x^{i}) \cap y^{i}|}{|h(x^{i})|}$$
(24)

$$RE = \frac{1}{p} \sum_{i=1}^{p} \frac{|h(x^{i}) \cap y^{i}|}{|y^{i}|}$$
(25)

$$F-score = \frac{\left(1+\beta^2\right) \cdot PR \cdot RE}{\beta^2 \cdot (PR+RE)}$$
(26)

To test the effectiveness of the algorithm, three sets of experiments were designed in this section to evaluate the performance of the sentiment analysis algorithm.

In the first group of experiments, the unary grammar features of words were used as the main features, and the characteristics of the combination of the unary grammar features with the adjacent sentences, the combination of the unary grammar features with the full text of the tweet features, and the combination of the unary grammar features with the adjacent sentences and the full text of the tweet features were applied to conduct the experiments. Based on the data characteristics and empirical *K*value selection (Bansal et al., 2022), K=5 was set as the benchmark, and a bidirectional test was conducted. According to the test results, K=5 and K=8 were selected as the parameter indicators in further experimentation.

In the second group of experiments, the features of MLkNN classifier select the combined features of unary grammar and binary grammar, and evaluate the situation of "unary + binary + adjacent sentences", "unary + binary + chapter" and "unary + binary + adjacent sentences + text". Finally, the adjacent sentence features and the full text of the tweet features are combined. The number of nearest neighbors K is 5 and 8, respectively. The results given in the table are the results when the average accuracy is iterated to convergence. This study labels the method with above procedure as S-MLkNN.

In the third group of experiments, the MLkNN classifier combining unary grammar features and binary grammar features is selected as the initial classifier of sentences, and the overall emotion of a tweet is classified by combining the adjacent sentences and text features. On this basis, the overall emotion label of a tweet is further modified through label correlation. Different nearest neighbor numbers K and different  $\alpha$  parameters of the training set are selected to obtain the corresponding emotion transfer matrix. This study labels the method with the above procedure as L-MLkNN.

# Results

**Emotion classification with basic MLkNN**. As shown in Table 2, the average accuracy could only reach about 42% if just unary grammar features are used in the MLkNN classifier. The use of the emotion transfer features of the adjacent sentences or the whole tweet could help a lot to improve the accuracy. When the value of K is 5, the accuracy rate is increased by about 13%, and the accuracy rate was further improved by 15% when the value of K is 8. Regretfully, the effect is not obvious compared with the classification results of the baseline system. It is mainly affected by the inaccurate initial classification, which has a relatively strong impact on the final classification result. This also indicates that when the classification accuracy of sentences is high, the accuracy of emotion classification could be better improved by considering the emotional transfer of the text through the characteristics of the adjacent sentences and the whole text.

*Emotion classification with S-MLkNN.* The second experiment proves that the binary grammar features could result in better initial classification results than the single unary grammar features as shown in Table 3. In this group of experiments, the selection of *K*-value, however, has little effect on the results.

*Emotion classification with L-MLkNN*. Table 4 shows the indicators after modifying the emotional label according to the relevance of the labels. The value of  $\alpha$  will affect the final emotion analysis result according to the combination of base classifiers and the number of neighbors. When the value of  $\alpha$  is 0.7, a higher accuracy could be obtained, and HL also reaches the lowest level in several groups of experiments. In addition, the value of *K* also impacts the results greatly. When *k* is 8, better HL and OE values could be achieved since more neighboring label sets are considered. However, the corresponding training time is longer and the training cost is higher.

# Discussion

This study compares the performance of the different algorithms on the Twitter dataset. If the number of predicted labels in the classification result of the classifier is more than two, only the emotional labels ranked in the top two are selected as the emotional labels of the text, that is, all texts are considered according to a maximum of two emotional labels. Table 5 shows the performance when the value of *K* in the MLkNN base classifier is 8 and the value of  $\alpha$  is 0.7. The results in Table 5 show that the improved L-MLkNN algorithm outperforms the other methods in the overall performance, where the advantage in RE is relatively more pronounced.

From all the experimental results, it is shown that the improved MLkNN algorithm can effectively boost the accuracy of emotion classification of short texts. After correction according to the relevance of labels, the classification effect of multi-labels in the sample set is significantly improved. However, most of the Twitter samples in the corpus are still singlelabeled. The limited proportion of multi-label samples in the training set leads to a small co-occurrence probability matrix, so the performance improvement effect of the classifier is not obvious for the corpus as a whole.

What is worth noticing is the performance of the three methods when classifying different texts. In order to better show the performance of the method on different lengths, the test set is grouped according to the length of the texts (the text number and classification result of each group is 8 and  $\alpha$  is 0.7). The corresponding F1 value is shown in Fig. 2. MLkNN is the base classifier classification result, S-MLkNN corrects the adjacent sentences and chapter features, and L-MLkNN corrects the classification result based on the label correlation. Figure 2 shows that the performance of the three methods varies little when the targeted

Table 2 Experimental results of emotion classification with basic MLkNN.							
	Features	SA	HL	OE	RL	AVP	
K = 5	Unary	0.4076	0.3249	0.7151	0.2888	0.4182	
	Unary + adjacent sentences	0.5359	0.3103	0.7006	0.2458	0.5365	
	Unary $+$ chapter	0.5878	0.3208	0.7190	0.2095	0.5589	
	Unary + adjacent sentences + text	0.6707	0.3070	0.6764	0.1726	0.6725	
K = 8	Unary	0.4208	0.2267	0.6826	0.1794	0.4219	
	Unary + adjacent sentences	0.5774	0.2198	0.7006	0.1758	0.5762	
	Unary $+$ chapter	0.5878	0.2209	0.6791	0.1923	0.5828	
	Unary + adjacent sentences + text	0.7179	0.2087	0.6803	0.1832	0.7205	

Table 3 Emotion analysis results of S-MLkNN.						
	Features	SA	HL	OE	RL	AVP
K = 5	Unary $+$ binary	0.4113	0.3454	0.6719	0.1755	0.4642
	Unary + binary + adjacent sentences	0.5439	0.3202	0.6411	0.1630	0.5434
	Unary + binary + chapter	0.6054	0.2848	0.6518	0.1547	0.5917
	Unary + binary + adjacent sentences + text	0.7029	0.2994	0.6219	0.1653	0.7046
K = 8	Unary + binary	0.6179	0.2318	0.6354	0.1967	0.6295
	Unary + binary + adjacent sentences	0.7689	0.2258	0.6012	0.1680	0.7578
	Unary + binary + chapter	0.7401	0.2292	0.5356	0.1794	0.7354
	Unary + binary + adjacent sentences + text	0.8103	0.1926	0.5459	0.1553	0.8056

Table 4 Emotion analysis results of L-MLkNN.							
		SA	HL	OE	RL	AVP	
$\alpha = 0.5$	K = 5	0.7872	0.2042	0.7701	0.2311	0.7884	
	K = 8	0.7926	0.1876	0.6174	0.2287	0.7913	
$\alpha = 0.7$	K = 5	0.8182	0.0809	0.7823	0.1526	0.7423	
	K = 8	0.8205	0.0917	0.5256	0.1598	0.8101	
$\alpha = 0.9$	K = 5	0.7658	0.1226	0.6023	0.0944	0.7643	
	K = 8	0.7766	0.1293	0.5198	0.1294	0.7701	

Table 5 Comparison of the performance of different algorithms.								
	AC	PR	RE	F-score				
ML-DT	0.5730	0.6842	0.6924	0.6883				
Rank-SVM	0.5539	0.6646	0.7165	0.7101				
CML	0.6247	0.7423	0.6348	0.6844				
MLkNN	0.5321	0.6559	0.7104	0.6655				
S-MLkNN	0.6012	0.8056	0.7324	0.7673				
L-MLkNN	0.5854	0.8101	0.8019	0.8060				



**Fig. 2** Performance of classification for texts with different amount words (F1 value).

texts are short texts. When classifying long texts, the performance of the improved algorithm is significantly higher compared with the base classifier.

The experimental verification on the Twitter corpus shows that on the same dataset, the method proposed in this chapter achieves better classification results than the traditional multilabel classifier. However, this method still highly relies on the labeling of the training set. From a machine-learning perspective, this is supervised learning. When the classification accuracy of the initial sentence is not high, the classification results might be disappointing. How to use smaller training samples and lower training costs to obtain efficient classifiers, and the performance of such methods on semi-supervised learning or even unsupervised learning are worth investigating further.

It has to be acknowledged that the small sample size of the dataset in this study may not be representative of the entire dataset. The model's ability to generalize to new data may be limited as a small dataset may not contain enough variation in the data, and potential sampling bias may exist. Applying the results of this research rashly to large data sets may not be robust enough when encountering outliers or unexpected data. In future studies, the scholars should pay more attention to balancing the efficiency of the model with smaller training sample sizes and the completeness of the model to cover various scenarios.

#### Conclusions

The process of computationally identifying and categorizing opinions expressed in a piece of text is important to provide a better understanding and services to online users. By considering not only individual in-sentence features but also the features in the adjacent sentences and the full text of the tweet, this study adjusts the MLkNN classifier to allow iterative corrections of the multi-label emotion classification and applies the new method to improve both the accuracy and speed of the emotion classification for short texts in Twitter. Except for the adjustments based on the emotion transfers, this study further takes the correlation between multiple emotion labels into consideration and iteratively updates the overall classification results. By carrying out three groups of experiments on the Twitter corpus, this study compares the performances of the base classifier of MLkNN, the sample-based MLkNN (S-MLkNN) and the label-based MLkNN(L-MLkNN). It is proven that the experiments offer the best performance when the value of K in the MLkNN base classifier is 8 and the value of  $\alpha$  is 0.7.

This study is an attempt to obtain an efficient classifier faster and more accurately. This method still works or even performs better for long texts. However, further work still needs to be carried out to improve semi-supervised learning or unsupervised learning algorithms, and to achieve better performance with smaller training samples and lower training costs.

### **Data availability**

The data that support the findings of this study are downloaded from www.kaggle.com. The data URLs: [https://www.kaggle.com/ datasets/kazanova/sentiment140].

Received: 9 January 2023; Accepted: 30 May 2023; Published online: 08 June 2023

#### Note

1 Sentiment140 Twitter corpus, https://www.kaggl`e.com/datasets/kazanova/ sentiment140, accessed on Aug 28, 2020.

#### References

- Ai Y, Chen Z, Wang S et al. (2018) Recognizing emotions in chinese text using dictionary and ensemble of classifiers. In: Third International Workshop on Pattern Recognition. SPIE, pp. 39–45
- Aman S, Szpakowicz S (2007) Identifying expressions of emotion in text. International Conference on Text, Speech and Dialogue. Springer, pp. 196–205
- Baeza-Yates R, Ribeiro-Neto B (1999) Modern information retrieval. ACM press, New York
- Bansal M, Goyal A, Choudhary A (2022) A comparative analysis of K-nearest neighbour, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning. Decis Analyt J 100071. https:// doi.org/10.1016/j.dajour.2022.100071
- Chen Y, Zhang Z (2018) Research on text sentiment analysis based on cnns and svm. In: 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA). IEEE, pp. 2731–2734
- Chou S, Hsu CL (2005) Mmdt: a multi-valued and multi-labeled decision tree classifier for data mining. Exp Syst Appl 28(4):799–812. https://doi.org/10. 1016/j.eswa.2004.12.035
- Dogan T, Uysal AK (2020) A novel term weighting scheme for text classification: Tf-mono. J Informet 14(4):101076. https://doi.org/10.1016/j.joi.2020. 101076
- Fei H, Ji D, Zhang Y et al. (2020) Topic-enhanced capsule network for multi-label emotion classification. IEEE/ACM Trans Audio Speech Lang Process 28:1839–1848. https://doi.org/10.1109/TASLP.2020.3001390
- Feng X, Hui K, Deng X et al. (2021) Understanding how the semantic features of contents influence the diffusion of government microblogs: Moderating role of content topics. Inf Manage 58(8):103547. https://doi.org/10.1016/j.im. 2021.103547
- Go A, Bhayani R, Huang L (2009) Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, p 12. 1(2009)
- Hu A, Flaxman S (2018) Multimodal sentiment analysis to explore the structure of emotions. In: KDD '18: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 350–358. https:// doi.org/10.1145/3219819.3219853
- Jiang X, Song C, Xu Y, Li Y, Peng Y (2022) Research on sentiment classification for netizens based on the BERT-BiLSTM-TextCNN model. PeerJ Comput Sci 8:e1005. https://doi.org/10.7717/peerj-cs.1005
- Keshtkar F, Inkpen D (2012) A hierarchical approach to mood classification in blogs. Nat Lang Eng 18(1):61–81. https://doi.org/10.1017/ \$1351324911000118
- Li X, Xie H, Rao Y, Chen Y, Liu X, Huang H, Wang F (2016) Weighted multi-label classification model for sentiment analysis of online news. In: 2016 International Conference on Big Data and Smart Computing (BigComp). https:// doi.org/10.1109/BIGCOMP.2016.7425916
- Liang J, Chen S, Jin Q (2019) Semi-supervised multimodal emotion recognition with improved wasserstein gans. In: 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). IEEE, pp. 695–703
- Liang Y, Tian T, Jin K, Yang X, Lv Y, Zhang X (2020) Robust text clustering with graph and textual adversarial learning. In: 2020 IEEE Fifth International Conference on Data Science in Cyberspace (DSC). https://doi.org/10.1109/ DSC50466.2020.00036
- Liao W, Ma Y, Cao Y et al. (2021) Two-stage finegrained text-level sentiment analysis based on syntactic rule matching and deep semantic. IEICE Trans Inf Syst 104(8):1274–1280. https://doi.org/10.1587/transinf.2020BDP0018
- Liu H, Guo H, Hu W (2021) Eeg-based emotion classification using joint adaptation networks. In: 2021 IEEE international symposium on circuits and systems (ISCAS). IEEE, pp. 1–5

- Liu S, Chen J (2015) A multi-label classification based approach for sentiment classification. Exp Syst Appl 42(3):1083–1093. https://doi.org/10.1016/j.eswa. 2014.08.036
- Lou Z, Wu Y, Fan C et al. (2020) Aspect-based sentiment analysis on convolution neural network and multi-hierarchical attention. In: 2020 International Conference on Technologies and Applications of Artificial Intelligence (TAAI). IEEE, pp. 193–198
- Ma C, Osherenko A, Prendinger H et al. (2005) A chat system based on emotion estimation from text and embodied conversational messengers. In: Proceedings of the 2005 International Conference on Active Media Technology, 2005 (AMT 2005). IEEE, pp. 546–548
- Onan A (2022) Bidirectional convolutional recurrent neural network architecture with group-wise enhancement mechanism for text sentiment classification. J King Saud Univ-Comput Inf Sci 34:2098–2117. https://doi.org/10.1016/j. jksuci.2022.02.025
- Paltoglou G, Thelwall M (2012) Twitter, myspace, digg: Unsupervised sentiment analysis in social media. ACM Trans Intelligent Syst Technol (TIST) 3(4):1–19. https://doi.org/10.1145/2337542.2337551
- Rajabi Z, Shehu A, Uzuner O (2020) A multi-channel bilstmcnn model for multilabel emotion classification of informal text. In: 2020 IEEE 14th International Conference on Semantic Computing (ICSC). IEEE, pp. 303–306
- Sadr H, Pedram MM, Teshnehlab M (2019) A robust sentiment analysis method based on sequential combination of convolutional and recursive neural networks. Neural Process Lett 50(3):2745–2761. https://doi.org/10.1007/s11063-019-10049-1
- Shu L, Xie J, Yang M et al. (2018) A review of emotion recognition using physiological signals. Sensors 18(7):2074. https://doi.org/10.3390/s18072074
- Sintsova V, Musat C, Pu P (2014) Semi-supervised method for multi-category emotion recognition in tweets. In: 2014 IEEE International Conference on Data Mining Workshop. IEEE, pp. 393–402
- Siriwardhana S, Kaluarachchi T, Billinghurst M et al. (2020) Multimodal emotion recognition with transformerbased self supervised feature fusion. IEEE Access 8:176274–176285. https://doi.org/10.1109/ACCESS.2020.3026823
- Taboada M, Brooke J, Tofiloski M et al. (2011) Lexicon-based methods for sentiment analysis. Comput Linguist 37(2):267–307. https://doi.org/10.1162/ COLI\_a\_00049
- Tang Y, Su J, Khan M (2021) Research on sentiment analysis of network forum based on bp neural network. Mob Netw Appl 26(1):174–183. https://doi.org/ 10.1007/s11036-020-01697-y
- Tasmin M (2018) Multi-dimensional aspect analysis of text input through human emotion and social factors. In: UbiComp '18: Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, pp. 1779–1781. https://doi.org/10.1145/3267305.3277817
- Ullah S, Talib M, Rana T et al. (2022) Deep learning and machine learning-based model for conversational sentiment classification. Comput Mater Continua 72(2):2323–2339. https://doi.org/10.32604/cmc.2022.025543
- Xie J, Chen B, Gu X, Liang F, Xu X (2019) Self-attention-based BiLSTM model for short text fine-grained sentiment classification. IEEE Access 2019 7:180558–180570. https://doi.org/10.1109/ACCESS.2019.2957510
- Yan D, Hu B, Qin J (2018) Sentiment analysis for microblog related to finance based on rules and classification. In: 2018 IEEE international conference on big data and smart computing (BigComp). IEEE, pp. 119–126
- Yang J, Jiang L, Wang C et al. (2014) Multi-label emotion classification for tweets in weibo: method and application. In: 2014 IEEE 26th International Conference on Tools with Artificial Intelligence (ICTAI). IEEE Computer Society, pp. 424–428
- Yang K, She C, Zhang W et al. (2019) Multi-label learning based on transfer learning and label correlation. CMC Comput Mater Continua 61(1):155–169. https://doi.org/10.32604/cmc.2019.05901

- Zhang M, Zhang K (2010) Multi-label learning by exploiting label dependency. In: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 999–1008
- Zhang M, Zhou Z (2006) Multilabel neural networks with applications to functional genomics and text categorization. IEEE Trans Knowl Data Eng 18(10):1338–1351. https://doi.org/10.1109/TKDE.2006.162
- Zhang M, Zhou Z (2007) MI-knn: a lazy learning approach to multi-label learning. Pattern Recognit 40(7):2038–2048. https://doi.org/10.1016/j.patcog.2006.12.019

#### **Author contributions**

Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work: conceptualization: XL and WZ; methodology: TS; software: ML; data acquisition: GZ; formal analysis: ZY and LY; data curation: TS and LY; validation: ML and ZY. Drafting the work or revising it critically for important intellectual content: writing—original draft preparation: LY, TS, and XL; writing—review and editing: XL and WZ; visualization: LY and TS. Final approval of the version to be published: project administration and supervision: WZ. All authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

#### **Competing interests**

The authors declare no competing interests.

#### Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

#### **Informed consent**

This article does not contain any studies with human participants performed by any of the authors.

#### **Additional information**

Correspondence and requests for materials should be addressed to Wenfeng Zheng.

Reprints and permission information is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/ licenses/by/4.0/.

© The Author(s) 2023