Open Text Classification Based on Dynamic Boundary Balance

Jianzhou Feng
Yanshan University
Ganlin Xu (xuglhb@163.com)
Yanshan University
Qikai Wei
Yanshan University

Research Article

Keywords: open text classification, two-stage learning model, decision boundary, open space risk, empirical risk

Posted Date: June 6th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1714048/v1

Open Text Classification Based on Dynamic Boundary Balance

Jianzhou Feng[1], Ganlin Xu[1*] and Qikai Wei[1]

[1]School of Information Science and Engineering, Yanshan University, Qinhuangdao, 066000, China.

*Corresponding author(s). E-mail(s): xuglhb@163.com;
Contributing authors: fjzwxh@ysu.edu.cn; mrweiqk@163.com;

Abstract

Traditional text classification tasks require the test set to contain the same classes (known/seen classes) as the training set. However, when the new classes (unknown/unseen classes) appear during testing, the traditional classifier cannot recognize them effectively. As a result, a qualified open text classifier should be able to classify known classes as well as identify unknown classes, e.i., open classfication. But current open text classifiers perform poorly on the task. To address the issue, we propose a two-stage learning model based on dynamic balance of decision boundary to complete the open text classification task. We call it open text classification with dynamic boundary balance (OCD2B). First, we construct a traditional classifier based on BERT(Bidirectional Encoder Representation from Transformers) model to classify the known classes. Second, the prior knowledge of known classes is used to dynamically determine the decision boundaries between known classes and unknown classes. We propose a boundary loss function as boundary balance strategy to simultaneously reduce open space risk and empirical risk. Experiments on two standard datasets show that our method achieves better performance than existing other methods. In particular, OCD2B has even greater advantages over the others especially when the proportion of unknown classes is large.

Keywords: open text classification, two-stage learning model, decision boundary, open space risk, empirical risk

1 Introduction

In the traditional supervised learning text classification task, the classifier requires the training set to have the same class space as the test set. [1] referred to it as closed-world assumption. This opinion has been widely adopted in many fields of natural language processing, such as emotion analysis, spam recognition, and news classification. However, the dynamic open environment frequently contains scenes that the model has never seen before. It is critical to distinguish these unknown scenes as much as possible from the known scenes. For example, in a dialogue system, user intent is typically very complex and some intents that the model has not learned always exist. In addition, when the e-commerce system classifies commodities using the commodity description, it also needs to identify the classes that are not recorded in the system. It is a challenging task because it is difficult to obtain prior knowledge of unknown classes for lack of unknown samples [2] and master the number of unknown classes during testing. In the field of natural language processing, a better classifier is required not only to classify known classes but also to obtain a novel mechanism to discover unknown classes. [1] referred to it as open (world) classification.

According to [3], open text classification is an m+1 classification task, where m is the number of known classes. l_i is the i^{th} known class label given a label set $L = \{l_1, l_2, ..., l_i, \ldots, l_m\}$ of a training set. The labels of all unknown classes are defined as l_{m+1}; thus, all classes that do not belong to L are labeled as unknown classes. Our task is to classify the m-class known classes into their corresponding classes correctly while identifying the $(m+1)^{th}$ class as suggested in [3–5], where the $(m+1)^{th}$ class represents the unknown class.

In recent years, some progress has been made in the research on open text classification. [1] proposed Center-Based Similarity (CBS) space learning method in [6] and the method assumed that each known class is wrapped in a "ball". The centroids of "ball" is obtained by averaging the vector, and the sphere acts as decision boundary (all radii are 0.5). According to [7], The BERT model is used to vectorize sentences, and then a post-processing method is proposed to learn adaptive decision boundary of "ball". Inspired by these works, we propose a two-stage model based on the dynamic balance of the decision boundary. To perform open text classification, the model utilize the prior knowledge of the known classes to dynamically adjust the decision boundary.

The main contributions are as follows:

1) We propose OCD2B, a novel approach for open text classification based on dynamic balance of decision boundary. The boundary loss function dynamically adjusts the decision threshold by the prior knowledge of known classes, and then obtain the decision boundary.

2) The experimental results on two standard datasets show that our method exhibits better performance than the previous methods.

The remainder of the paper is organized as follow. Section 2 introduces related works as well as background information on open classification. The method are also introduced in this section. Section 3 presents the details of the

method, including the architecture of classification net, the choice of decision boundary and the mechanism of open text classification. Section 4 provides two experiments to evaluate the method compared with other models and demonstrate its effectivenes. Finally, Section 5 presents this paper's conclusion.

2 Related Work

There are some previous methods for open classification. [8] initially propose the concept of open space risk in computer vision to evaluate open classification. They recognize unknown images using the SVM's hyperplane of binary classification. The concept of open space risk was subsequently applied to the field of natural language processing. [4] fit the probability distributions for each class based on statistical Extreme Value Theory (EVT) and use a Weibull-calibrated multiclass SVM for open classification task. [9] build a Weibull-calibrated SVM classifier using EVT that further improves the performance. They referred it as Compact Abating Probability (CAP) model. However, these methods determine decision boundaries by using unknown class samples. [1] reduce the open space risk by setting the fixed boundary of each known class in sphere. However, traditional machine learning [1, 8] only focused on the n-gram information among sentences instead of high dimensional semantic features. Results show that the performance of these methods is poorer than deep learning method according to [10].

Because machine learning is limited in capturing high-level semantic features, many researches employ deep neural network to solve open classification task. [11] propose a new method called OpenMax in computer version for open set recognition and one weak assumption is that it still need prior knowledge of unknown classes. [10] extract text features by Convolutional Neural Network (CNN), and then the distance between the sentence and the class center was mapped into a probability value through the Weibull distribution. [12] considered the highest probability score of the final softmax of the classifier and compared it with 0.5 to indicate whether it belongs to known classes or unknown classes. The above two methods cannot achieve good results in open classification due to fixed thresholds, which are similar to [1]. [3] replaced softmax with sigmoid in the last layer of CNN, and a threshold is determined for each known class as the decision boundary based on statistics. Instead, CNN often focuses on to the local information of the text, which assuming that the output probability of the training set conforming to the Gaussian distribution frequently fails in large amounts of data. [2] utilize margin loss to increase the inter-class variance and reduce the intra-class variance for unknown intent detection. Then density-based detection algorithm called local outlier factor LOF is used to detect unknown classes. However, it does not take open space into consideration for distinguishing the open intent. [13] propose a post-processing method to extract text features using neural network to discover the unknown intention of dialogue system. [7] utilize BERT model to extract text feature, and then propose a post-processing method to reduce

open space risk and empirical risk. [31] takes advantage of the sequence-to-sequence language model BART to creat distributionally shifted example from the training examples, which aimed at learning an open representation. [33] focus on the unsupervised out-of-domain detection. They propose a supervised contrastive learning objective to minimize intra-class variance and maximize inter-class variance. [32] start from the nature of out-of-domain intent classification and futher utilize K-Nearest Neighbors of in-domain intents to obtain discriminative semantic features for out-of-domain detection.

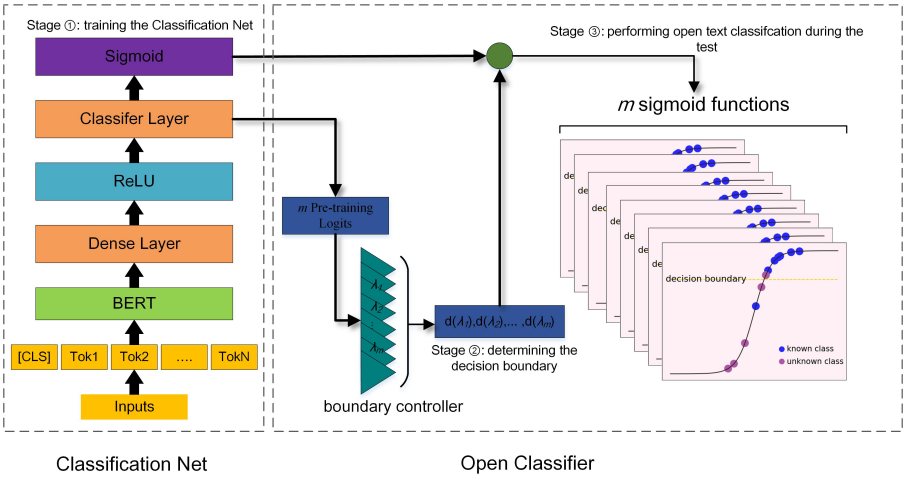

Fig. 1 Model architecture of OCD2B.The classification net is a traditional classifier built by BERT, which is used to classify known classes and provide prior knowledge of known classes for decision boundaries. The open classifier utilizes the prior knowledge of known classes and dynamically adjust the decision boundary, which aimed to distinguish the unknown classes. The classification net and open classifier work together to perform open text classification.

3 Method

The structure of OCD2B is shown in Fig. 1 and the entire model consists of two parts. The classification net classifies known classes, and the open classifier can identify unknown classes using a dynamic boundary balance strategy. These components work together to complete the open text classification.

3.1 Classification Net

A traditional classifier(classification net) is constructed based on the BERT model, and it used to classify known classes and provide the prior knowledge of known classes for decision boundaries. The architecture of the classifier is shown in the left of Fig. 1, we train the classification net in the stage ①. The token $[CLS, Tok_1, Tok_2, ..., Tok_n]$ is input into the classification net, and then output token embedding $[C, T_1, Tok_2, \ldots, Tok_Q] \in \mathrm{R}^{(Q+1)\times H}$. As suggested

in [7], we perform mean-pooling on these token embeddings to obtain averaged vector $\boldsymbol{x}_i$:

$$x_i = mean - pooling([C, T_1, T_2, .., T_Q]) \tag{1}$$

where Q is the length of the sequence, H is the size of the hidden layer.

Then, the vector $\boldsymbol{x}_i$ is used to obtain the m probability values to perform the classification by two linear layers. The first linear layer is dense layer followed by a ReLU function as the activation function. The second linear layer called classifier layer plays the role of dimension reduction, reducing the high-dimensional vector obtained by the dense layer to m dimension, which using sigmoid function as the activation function to output m probability values $p_{1:m}$.

$$p_{1:m} = \sigma(W_2(ReLU(W_{1:x_{1:k}} + b_1)) + b_2) \tag{2}$$

where $p_{1:m}$ is the probability sequence of one test sample corresponding to each known class. W_1, b_1, W_2 and b_2 are trainable weights. σ is the sigmoid function. We select the binary cross-entropy loss function $\mathcal{L}_c$, as follows:

$$\mathcal{L}_c = -\sum_{i=1}^{n}\sum_{j=1}^{m} I(t_i = l_i)log(t_i = l_i)) + (1 - I(t_i \neq l_i))log(1 - (t_i = l_i)) \tag{3}$$

where n is batch size, p$(\cdot)$ is the probability value of sigmoid output, t_i is the real sample label, l_i is the expected sample label, and I$(\cdot)$ is defined as follows:

$$I(\cdot) = \begin{cases} 1, \ t_i = l_i \\ 0, \ t_i \neq l_i \end{cases} \tag{4}$$

As shown in Fig. 1, after the training classification net, we extract the m dimension pre-training logits from classifer layer as the prior knowledge of known classes to determine the decision boundaries.

3.2 Dynamic Boundary Balance

The right of Fig. 1 is the open classifier. Its main function is to execute the boundary balance strategy based on classification net and obtain the decision threshold through training. The BERT model has strong text feature representation capabilities, so it can maximize inter-class variance and minimize intra-class variance. In the traditional text classification, the sigmoid probability scores of known classes are often in the "higher" part of the sigmoid function. However, because unknown classes do not participate in the training of the model, the output probability score often falls on the "lower" part. Therefore, with classification net "dividing" known classes and unknown classes in different space, open text classification can be implemented by introducing a dividing line as the decision boundary on the sigmoid function.

In the phase, m dimension pre-training logits is input into the boundary controller to determine decision boundaries d(λ_1), d(λ_2),..., d(λ_m) in the right of Fig. 1. Specifically, we propose a boundary loss function $\mathcal{L}_\lambda$ dynamically

adjust the decision λ by the prior knowledge of known classes in the stage ②, and then obtain the decision boundary to segment known classes and unknown classes, as follows:

$$\mathcal{L}_\lambda = \sum_{i=1}^{n} k\delta(y_i - \lambda_{y_i}) + (1-\delta)(\lambda_{y_i} - y_i) \tag{5}$$

where y_i is the output of train set at the classifier layer, λ_{y_i} denotes the decision threshold of the known class corresponding to y_i, and n is batch size. During the initialization of λ_{y_i}, its range is within (-∞, +∞), conforming to the Gaussian distribution. δ is defined as follows:

$$\delta = \begin{cases} 1, & y_i \geq \lambda_{y_i} \\ 0, & y_i < \lambda_{y_i} \end{cases} \tag{6}$$

During the training of the open classifier, λ is updated by the following:

$$\lambda = \lambda - \eta \frac{\partial \mathcal{L}_\lambda}{\lambda} \tag{7}$$

where η is the learning rate. we dynamically select λ by Algorithm 1.

Algorithm 1 Decision Threshold Selection Algorithm

Require: initial decision threshold $\lambda_{1:m}$.
Ensure: final decision threshold $\grave{\lambda}_{1:m}$.

```
while minimize L_λ do
    compute boundary loss L_λ by Eq. (5) during training the open classifier.
    update λ̀_{1:m} ← λ_{1:m} by Eq. (7).
end while
return λ̀_{1:m}
```

The decision boundary can balance open space risk and empirical risk to perform open classification. We assume that if $(y_i\text{-}\lambda_i) \geq 0$, the known samples will be below the corresponding decision boundaries and identified as unknown classes, which will cause the empirical risk. However, if $(\lambda_i\text{-}y_i) < 0$, more unknown samples will be above the corresponding decision boundaries and identified as known classes. So we need "move up" the decision boundaries to reduce the open space risk. We define k as a balance factor to make decision boundaries adaptive to known class space. The left loss $(y_i\text{-}\lambda_i)$ of Eq. (5) will increase with the increase in k and reduce the open space risk. For example, as shown in Fig. 2, when $k = 1$, a large number of known class samples are under the decision boundary, which affects the classification performance of the model. When $k = 4$, the decision boundary will "move down", and it can better distinguish known classes and unknown classes. However, If the k is too large, more unknown classes will be identified as known classes, which leading

to empirical risk. Thus, we utilize a grid search algorithm for selecting k and our method achieves the best results when $k = 14$.

During training, the "boundary loss" is calculated repeatedly through Eq. (5) to minimize $\mathcal{L}_\lambda$. Finally, the decision boundary can achieve balance between known and unknown classes. it can reduce the open space risk and empirical risk at the same time.

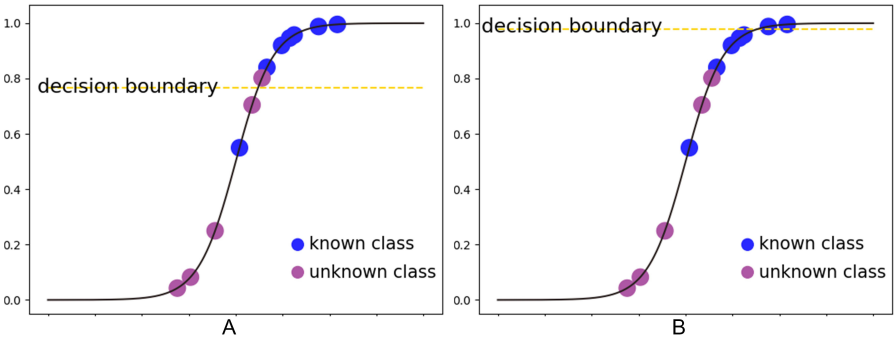

Fig. 2 (A) shows $k = 1$, and (B) shows $k = 4$. Blue points represent unknown classes, and purple points represent known classes

3.3 Open Classification

After the two stages, each known class has a definite decision threshold λ. Each decision threshold λ determines a unique decision boundary d(λ). In the stage ③, the output probability of the unknown classes becomes below the corresponding decision boundary, and the known classes becomes above the decision boundary. If the predicted probability value of a test sample output at classification net is less than the decision boundary of the corresponding known class, then it belongs to the unknown class. Otherwise, it belongs to the known class with the highest probability value. The formula for implementing open classification is as follows:

$$\hat{y} = \begin{cases} unknown\ class, & if\ d(y_i) < d(\lambda_{y_i}) \\ argmax_{l_i \in L} d(\lambda_{y_i}), & otherwise \end{cases} \tag{8}$$

where L is the set of labels for known classes, m is the number of known classes, and d($\cdot$) is the sigmoid function.

The sigmoid activation function has three advantages. First, it can make the range of decision boundary be within (0, 1). Second, it is totally differentiable with different λ. Finally, it can introduce nonlinear factors and be convenient for derivation.

4 Expriment

In this section, two datasets are firstly introduced. Then we introduce six baselines. Next, in the experiment of section 4.3, we introduce the evaluation metrics and in the experiment of section 4.4, we analyzed the experimental results on two datasets. Finally, we make a fine-grained analysis of the experimental results

4.1 Datasets

Two datasets are used to better compare with other models to complete the experiments. Both datasets belong to the short text dataset because the maximum token length of BERT is 512.

OOS It is a dataset of semantic intention classification, covering common intent in daily life [14]. It contains 150 classes. Each class consists of 150 labeled sentences. We select 100 sentences as the training set, 20 sentences as the validation set and 30 sentences as the test set for each class. The maximum token length in the dataset is 28, and the average token length is 8.31.

BANKING It is a fine-grained problem consultation dataset in the banking field [15]. It contains 77 classes, and the number of queries in each class is different, with a total of 13083 queries. We select 9003 queries as the training set, 1000 queries as the validation set, and 3080 queries as the test set. The maximum token length of the queries is 79, and the average token length is 11.91.

4.2 Baselines

Our OCD2B compared with the following baseline model and the experimental results are shown in Tables 1 and 2, respectively.

MSP The model utilizes the probability score by softmax as the classification basis after the last linear layer, and then selects the class with the highest probability for comparison at 0.5 [12]. If it is lower than 0.5, then it is judged as unknown class; otherwise, it belongs to one of the known classes.

DOC The algorithm replaces softmax with sigmoid and determines the threshold for each known class based on statistical method as the decision boundary of each class to find unknown classes [3].

OpenMax OpenMax is an open set recognition by CNNs with a softmax output layer in computer vision, we adapt it for open text classification [11]. Firstly, it uses logits as the feature space and fit a Weibull distribution. Then, it recalibrate the confidence scores with the OpenMax Layer to perform open text classification.

DeepUnk It uses margin loss to increase inter-class variance and reduce intra-class variance, and then uses the outlier detection algorithm LOF to find new classes [2].

ADB The BERT model is used to extract text features, and the mean value of the vector is used as the class center [7]. A post-processing method of defined

loss function is proposed to determine the decision boundary. Finally, it carries out open classification by calculating the distance between data points and each class.

OCD2B ($k = 1$) It is the basic OCD2B ($k = 1$). We compare it with conventional OCD2B ($k = 14$) to show the effect of balance factor $k = 1$ for the select of decision boundaries.

Table 1 Accuracy and macro-F1 score of open classification with different known class proportions on OOS

Methods	25%		50%		75%	
	Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score
MSP	47.02	47.62	62.96	70.41	74.07	82.38
DOC	74.97	66.37	77.16	78.26	78.73	83.59
OpenMax	68.50	61.99	80.11	80.56	76.80	73.16
DeepUnk	81.43	71.16	83.35	82.16	83.71	86.23
ADB	87.59	77.19	86.54	85.05	**86.32**	**88.53**
OCD2B ($k = 1$)	90.27	72.49	84.10	74.53	73.33	69.29
OCD2B	**91.97**	**81.63**	**88.86**	**85.42**	85.94	86.76

Table 2 Accuracy and macro-F1 score of open classification with different known class proportions on BANKING

Methods	25%		50%		75%	
	Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score
MSP	43.67	50.09	59.73	71.18	75.89	83.60
DOC	56.99	58.03	64.81	73.12	76.77	83.34
OpenMax	49.94	54.14	65.31	74.24	77.45	84.07
DeepUnk	64.21	61.36	65.31	74.24	78.53	84.31
ADB	78.85	71.62	78.86	80.90	81.08	85.96
OCD2B ($k = 1$)	**85.89**	65.98	79.20	76.96	69.38	73.59
OCD2B	80.81	**73.11**	**80.31**	**81.65**	**81.52**	**86.20**

4.3 Evaluation Metrics

We hold that open text classification is an extension of multi-text classification. Previous works conducted by [3] and [2] take macro-F1 score as the only evaluation metric. In this paper, we use macro-F1 score and accuracy as evaluation metrics. In the experiment, This study varies the number of training classes and use 25%, 50%, and 75% classes for training and all classes for testing. Taking OOS as an example, for 25% classes, we use 38 classes for training and all 150 classes for testing. We count 10 times and average the results in every setting. The experimental results are shown in Tables 1 and 2, and the best results are highlighted in bold.

4.4 Experimental Settings

We use the BERT model (BERT-uncased, with 12-layer transformer, 768 hidden size, and 12 self-attention heads) implemented by Pytorch. The learning rate of classifica-tion net is 2e-5, and the learning rate of open classifier is 0.05. The batch size during the training is 64, and the batch size during evaluation and testing is 32. We use Adam as optimizer in classification net and open classifier.

Table 3 Accuracy of classification net perfroming traditonal classfication on validation set and test set, respectively. Traditional classification refer to test set and training set have the same class space

Methods	25%		50%		75%	
	OOS	BANKING	OOS	BANKING	OOS	BANKING
Validation set	98.55	98.88	97.87	96.09	95.94	94.30
Test set	97.69	95.49	96.75	93.63	95.78	92.04

4.5 Result Analysis

The results of BANKING and OOS are shown in Tables 1 and 2, respectively. The tables show the following ob-servations:
1. The performance of OCD2B is better than MSP, DOC, and DeepUnk at all settings. It still has advantages in some respects compared with ADB.
2. For 25% setting, our method achieves quite good results. The accuracy and f1-score is ahead of other models on OOS. On BANKING, the f1-score is higher than other baselines but its accuracy(80.81) is less than OCD2B(k = 1)(85.89). We analyze that the decision boundaries of OCD2B(k = 1) are in "higher" position so that some known classes are incorrecty classifiied as unknown classes but a large number of unknown classes is correctly recognized. As a consequence, the accuracy score is very high but the f1-score is very low.
3. For 50% settings, our model is significantly ahead of others. Compared with the best results of other baselines, our method improves accuracy on OOS by 2.32%, on BANKING by 1.45% and improves f1-score on OOS by 0.37%, on BANKING by 0.75%. We hold that the selection of decision boundaries plays an important role. The method of dynamically determining the threshold based on the prior knowledge of known classes acts as a good decision boundary. It considerably reduces the inaccurate classification, for example, one known class are incorrectly identified as other known classes or unknown classes are incorrectly identified as known classes.
4. For 75% settings, our method still outperforms MSP, DOC, and DeepUnk. However, OCD2B are slightly worse than those of ADB on OOS but outperforms it on BANKING. We analyze the reason is that the traditional classification performance of classification net decline with the increase of

known classes as shown in Table 3. That is, some known classes will be classified as other known classes.

Secondly, we also perform fine-grained expriments as suggested in Zhang,

Table 4 Results of open classification with different known class proportions on OOS. "Known" and "Open" denote the macro F1-score over known classes and open class respectivel

Methods	25%		50%		75%	
	Known	Open	Known	Open	Known	Open
MSP	47.53	50.88	70.58	57.62	82.59	59.08
DOC	65.96	81.98	78.25	79.00	83.69	72.87
OpenMax	61.62	75.76	80.54	81.89	73.13	76.35
DeepUnk	70.73	87.33	82.11	85.85	86.27	81.15
ADB	76.80	91.84	85.00	88.65	**88.58**	83.92
OCD2B ($k = 1$)	71.92	94.07	74.35	88.14	69.24	75.24
OCD2B	**81.44**	**94.93**	**85.35**	**91.07**	86.68	**84.61**

Table 5 Results of open classification with different known class proportions on BANKING. "Known" and "Open" denote the macro F1-score over known classes and open class respectively

Methods	25%		50%		75%	
	Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score
MSP	50.55	41.43	71.97	41.19	84.36	39.23
DOC	57.85	61.42	73.59	55.14	83.91	50.60
OpenMax	54.28	51.32	74.76	54.33	84.64	50.85
DeepUnk	60.88	70.44	77.74	69.53	84.75	58.54
ADB	70.94	84.56	80.96	78.44	86.29	66.47
OCD2B ($k = 1$)	64.66	**91.18**	76.85	**81.04**	73.83	60.01
OCD2B	**72.20**	85.90	**81.76**	80.26	**86.16**	**68.62**

Xu, and Lin (2021). Tables 4 and 5 show the macro F1-score on open intent and known intents respectively. we can observe that our method still achieves better performance in most settings compared with other models. In the 75% setting on OOS, our model(86.68) is slightly worse than but closed to ADB on known classes(88.58). In addtion, the basic OCD2B($k = 1$) achieves best score in the 25% and 50% settings on open class of BANKING. That is because the decision boundaries are "higher" position the same as 25% setting of open classification.

Finally, as shown in Fig. 3, we record the training loss, accuracy and f1-score of open classifier during training for 50% known classes on BANKING. We can observe that our method enable decision boundaries converge quickly by approximately 6 epochs.

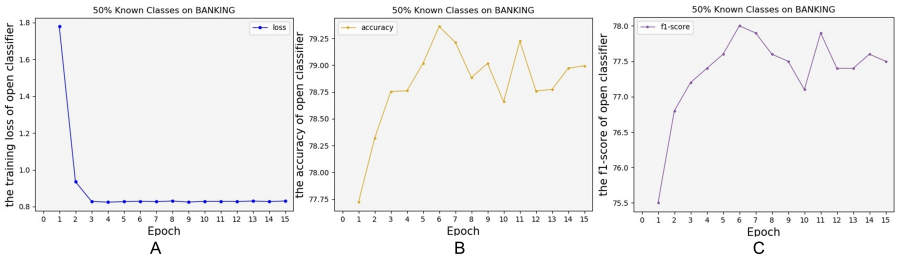

Fig. 3 (A), (B) and (C) shows the training loss, accuracy and f1-score of open classifier during training, respectively

5 Conclusion

In this paper, we propose a new method called OCD2B for open text classification. First, we utilize the BERT to construct a traditional classifier classification net that can effectively separate known classes from unknown classes. Then, on this basis, we utilize the trained threshold to obtain the decision boundary to perform open text classification. Finally, we showed that OCD2B performs better than the state-of-the-art methods from both the text classification domains.

Subsequently, we plan to continue to improve the accuracy of traditional classification when maintaining the ability to "separate" known classes from unknown classes, such as using xLNet model. We also consider the meta learning mechanism. The ability of lifelong learning is added on the basis of our model. When sufficient unknown classes of a certain class are learned, the model adds it to the known class set to have the learning ability of "people". More importantly, we will continue to explore the classification of unknown samples that belongs to an unknown class.

Declarations

Ethics approval and consent to participate. Not applicable.

Consent for publication. Yes.

Availability of data and materials. The datasets generated during the current study are available from the corresponding author on reasonable request.

Competing interests. The authors have no competing interests to declare that are relevant to the content of this article.

Funding. This work is supported by the National Natural Science Foundation of China (62172352), the Nature Scientist Fundation of Hebei Province (F2019203157), Scientific and technological research projects of colleges and universities in Hebei Province (ZD2019004) and sub-project of the National Key Research and Development Program (2020YFC0833404).

Authors' contributions. X.G. conducted experiments and wrote the initial draft. F.J. provided study materials and reviewed the draft. W.Q. checked the final draft. All authors contributed to the article and approved the submitted version.

Acknowledgements. This work is supported by the National Natural Science Foundation of China (62172352), the Nature Scientist Fundation of Hebei Province (F2019203157), Scientific and technological research projects of colleges and universities in Hebei Province (ZD2019004) and sub-project of the National Key Research and Development Program (2020YFC0833404).

References

[1] Fei, G., Liu, B.: Breaking the closed world assumption in text classification. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 506–514 (2016)

[2] Lin, T.-E., Xu, H.: Deep unknown intent detection with margin loss. arXiv preprint arXiv:1906.00434 (2019)

[3] Shu, L., Xu, H., Liu, B.: Doc: Deep open classification of text documents. arXiv preprint arXiv:1709.08716 (2017)

[4] Scheirer, W.J., Jain, L.P., Boult, T.E.: Probability models for open set recognition. IEEE transactions on pattern analysis and machine intelligence **36**(11), 2317–2324 (2014)

[5] Xu, H., Liu, B., Shu, L., Yu, P.: Open-world learning and application to product classification. In: The World Wide Web Conference, pp. 3413–3419 (2019)

[6] Fei, G., Liu, B.: Social media text classification under negative covariate shift. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 2347–2356 (2015)

[7] Zhang, H., Xu, H., Lin, T.-E.: Deep open intent classification with adaptive decision boundary. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 14374–14382 (2021)

[8] Scheirer, W.J., de Rezende Rocha, A., Sapkota, A., Boult, T.E.: Toward open set recognition. IEEE transactions on pattern analysis and machine intelligence **35**(7), 1757–1772 (2012)

[9] Jain, L.P., Scheirer, W.J., Boult, T.E.: Multi-class open set recognition using probability of inclusion. In: European Conference on Computer Vision, pp. 393–409 (2014). Springer

[10] Neal, L., Olson, M., Fern, X., Wong, W.-K., Li, F.: Open set learning with counterfactual images. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 613–628 (2018)

[11] Bendale, A., Boult, T.E.: Towards open set deep networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1563–1572 (2016)

[12] Hendrycks, D., Gimpel, K.: A baseline for detecting misclassified and out-of-distribution examples in neural networks. arXiv preprint arXiv:1610.02136 (2016)

[13] Lin, T.-E., Xu, H.: A post-processing method for detecting unknown intent of dialogue system via pre-trained deep neural network classifier. Knowledge-Based Systems **186**, 104979 (2019)

[14] Larson, S., Mahendran, A., Peper, J.J., Clarke, C., Lee, A., Hill, P., Kummerfeld, J.K., Leach, K., Laurenzano, M.A., Tang, L., et al.: An evaluation dataset for intent classification and out-of-scope prediction. arXiv preprint arXiv:1909.02027 (2019)

[15] Casanueva, I., Temčinas, T., Gerz, D., Henderson, M., Vulić, I.: Efficient intent detection with dual sentence encoders. arXiv preprint arXiv:2003.04807 (2020)

[16] Bendale, A., Boult, T.: Towards open world recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1893–1902 (2015)

[17] Brychcín, T., Král, P.: Unsupervised dialogue act induction using gaussian mixtures. arXiv preprint arXiv:1612.06572 (2016)

[18] Caron, M., Bojanowski, P., Joulin, A., Douze, M.: Deep clustering for unsupervised learning of visual features. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 132–149 (2018)

[19] Doan, T., Kalita, J.: Overcoming the challenge for text classification in the open world. In: 2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC), pp. 1–7 (2017). IEEE

[20] Choi, D., Shin, M.C., Kim, E., Shin, D.R.: Outflip: Generating out-of-domain samples for unknown intent detection with natural language attack. arXiv preprint arXiv:2105.05601 (2021)

[21] Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)

[22] Gangal, V., Arora, A., Einolghozati, A., Gupta, S.: Likelihood ratios and generative classifiers for unsupervised out-of-domain detection in task oriented dialog. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 7764–7771 (2020)

[23] Chen, D., Yu, Z.: Gold: improving out-of-scope detection in dialogues using data augmentation. arXiv preprint arXiv:2109.03079 (2021)

[24] Hsu, Y.-C., Lv, Z., Schlosser, J., Odom, P., Kira, Z.: Multi-class classification without multi-class labels. arXiv preprint arXiv:1901.00544 (2019)

[25] Kim, J.-K., Kim, Y.-B.: Joint learning of domain classification and out-of-domain detection with dynamic class weighting for satisficing false acceptance rates. arXiv preprint arXiv:1807.00072 (2018)

[26] Qin, L., Che, W., Li, Y., Wen, H., Liu, T.: A stack-propagation framework with token-level intent detection for spoken language understanding. arXiv preprint arXiv:1909.02188 (2019)

[27] Rozsa, A., Günther, M., Boult, T.E.: Adversarial robustness: Softmax versus openmax. arXiv preprint arXiv:1708.01697 (2017)

[28] Hughes, M., Li, I., Kotoulas, S., Suzumura, T.: Medical text classification using convolutional neural networks. In: Informatics for Health: Connected Citizen-Led Wellness and Population Health, pp. 246–250. IOS Press, ??? (2017)

[29] Zhang, H., Xu, H., Lin, T.-E., Lyu, R.: Discovering new intents with deep aligned clustering. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 14365–14373 (2021)

[30] Zhan, X., Xie, J., Liu, Z., Ong, Y.-S., Loy, C.C.: Online deep clustering for unsupervised representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6688–6697 (2020)

[31] Shu, L., Benajiba, Y., Mansour, S., Zhang, Y.: Odist: Open world classification via distributionally shifted instances. In: Findings of the Association for Computational Linguistics: EMNLP 2021, pp. 3751–3756 (2021)

[32] Zhou, Y., Liu, P., Qiu, X.: Knn-contrastive learning for out-of-domain intent classification. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 5129–5141 (2022)

[33] Zeng, Z., He, K., Yan, Y., Liu, Z., Wu, Y., Xu, H., Jiang, H., Xu, W.: Modeling discriminative representations for out-of-domain detection with supervised contrastive learning. arXiv preprint arXiv:2105.14289 (2021)