Reducing Risks In Open Set Recognition

Xueyang Yu
School of Information Science and Technology
ShanghaiTech University
yuxy1@shanghaitech.edu.cn
2020533050

Yijie Fan
School of Information Science and Technology
ShanghaiTech University
fanyj@shanghaitech.edu.cn
2020533120

Xinchen Jin
School of Information Science and Technology
ShanghaiTech University
jinxch@shanghaitech.edu.cn
2020533045

Abstract

The ability to identify whether or not a test sample belongs to one of the semantic classes in a classifier's training set is essential to practical deployment of the model. This task is termed open-set recognition (OSR) and has received significant attention in recent years. In this paper, we first modified several traditional machine learning and deep learning methods that are designed to solve closed set problems (including SVM, Logistic Regression, NN and neural network), and apply them on OSR problems. However, these models focus mainly on known classes, which may not be efficient for distinguishing the unknown classes. A natural thought to reduce risks is to represent unknown classes while training. One of the best algorithms is Adversarial Reciprocal Point Learning (ARPL). In this learning framework, reciprocal point and adversarial margin constraint are proposed to reduce the open space risk. We implemented this model and the results showed a huge boost on recognizing unknown classes compared with previous traditional ML algorithms. We further tested the model with larger datasets to prove its robustness. Last, we suggested some new aspects to work on in the future.

1 Introduction

Classic image classification problem usually assumes that data is a close set, i.e., categories appeared in testing set should all be covered by training set.However, in real-world applications, such as facial recognition and automatic driving, samples of unseen classes may appear in testing phase, misclassification of these unknown samples may lead to catastrophic results. To break the limitations of close set, Open Set Recognition (OSR) [1] was proposed, which has two sub-goals: known class classification and unknown class detection. In below, We formalize the problem of OSR and highlight its differences with *closed-set recognition*.

Open-set recognition. First, consider a labelled training set for a classifier $\mathcal{D}_{\text{train}} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N} \subset \mathcal{X} \times \mathcal{C}$. Here, $\mathcal{X}$ is the input space and $\mathcal{C}$ is the set of 'known' classes. In the closed-set scenario, the model is evaluated on a testing set in which the labels are also drawn from the same set of classes, *ie*, $\mathcal{D}_{\text{test-closed}} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{M} \subset \mathcal{X} \times \mathcal{C}$. In the closed-set setting, the model returns a distribution over the known classes as $p(y|\mathbf{x})$. Conversely, in OSR, test images may also come from unseen classes $\mathcal{U}$, giving $\mathcal{D}_{\text{test-open}} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{M'} \subset \mathcal{X} \times (\mathcal{C} \cup \mathcal{U})$. In the open-set setting, in addition to returning the distribution $p(y|\mathbf{x}, y \in \mathcal{C})$ over known classes, the model also returns a score $\mathcal{S}(y \in \mathcal{C}|\mathbf{x})$ to indicate whether or not the test sample belongs to *any* of the known classes. [2]

In this project, we first modify some close set algorithms, including SVM, NN algorithm, Logistic Regression and Neural Network, to test their performance on open set problems. We train those models with selected classes and use threshold to detect unknown samples during testing. Those models generally have some misclassifications on the simplest dataset MNIST.

The reason accounting for the result is that we didn't take unknown features into consideration while training. To be more specific, the question we want to solve is *How to identify a dog?*, while the traditional classification methods aim to learn *"what is a dog"* so that when training, the model will only see a limited part of the whole open space. Therefore when identifying unknown objects, the uncertainty greatly increases.

A natural thought is to learn how to represent unknown classes. A novel learning framework termed *Adversarial Reciprocal Point Learning (ARPL)* is proposed to solve OSR problems.[3] the framework is mainly composed of two parts. First, for each known class, a *Reciprocal Point* is learned to model their latent open space in the feature space. Then, adversarial margin constraint among multiple known categories is used to bound the unknown classes space and reduce risks. We implement this method and results show much better performances on both known classes classification accuracy and unknown classes detection accuracy. Further, we test this framework on larger dataset such as CIFAR10 and CIFAR100-50, the results are all satisfying.

2 Methodology

2.1 Logistic Regression

Logistic Regression can be viewed as a generalized linear model using the sigmoid function to train a model for binary classification. In the training process,we use the gradient descent method to minimize the loss function:

$$\mathcal{L}(\hat{y}, y) = \sum_{i=1}^{m} - \left(y^{(i)} \log(\hat{y}^{(i)}) + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)}) \right) \tag{1}$$

For multi-classification, we create one model for each class, this method is called One-vs-all classification method. After calculating the probability of the input with each model, the model with the highest probability is used to classify the input. And for the open set recognition task, before the model gives its final result, we calculate the value of the score:

$$Z = w^T x + b$$

if the score is smaller than a hyper-parameter θ where $0 < \theta < 2$,then the result will be classified as unknown. Otherwise, the model will classify the input with the same label of the highest probability label.

2.2 Nearest neighbors distance ratio open-set classifier

The Nearest neighbors distance ratio open-set classifier first calculates the distance between the test data point s and every training data point in the feature space and obtains the nearest neighbor t and the second nearest neighbor u such that $\theta(t) \neq \theta(u)$ where $\theta(x) \in \mathcal{L} = \{l_1, l_2, \ldots, l_n\}$ represents the class of a sample x and $\mathcal{L}$ is the set of known labels. Then we calculate the distance ration:

$$R = d(s, t)/d(s, u)$$

in which $d(x_1, x_2)$ is the distance between point x_1 and x_2 in the feature space. If R is less than or equal to the specified threshold T where $0 < T < 1$, s will be classified with the same label of t. Otherwise, it will be classified as unknown. [4]

2.3 SVM

For the open-set recognition problem, we train our SVM using the Gaussian radial basis function kernel. We first generate multiclass probability [5] and then set a hyperparameter P_{thres}. If $maxP_i < P_{thres}$, we choose to label the sample as unknown. Otherwise we label the sample as $\underset{i}{\operatorname{argmax}} P_i$.

2.4 CNN

We train a simple CNN architecture with two hidden layers using ReLU activation and max pooling. Then we set a hyper parameter P_{thres}. If $maxP_i < P_{thres}$, we choose to label the sample as unknown. Otherwise we label the sample as $\arg\max_i P_i$.

2.5 Adversarial Reciprocal Points Learning

Reciprocal Points The *reciprocal point* $\mathcal{P}^k$ of category k is regarded as the latent representation of the all data points that don't belong to category k. Hence, the samples of other parts in open space should be closer to the reciprocal point $\mathcal{P}^k$ than the samples of category k.
Specifically, Given an deep embedding function $\mathcal{C}$ with learnable parameters θ, sample x and reciprocal point $\mathcal{P}_k$, their distance $d(\mathcal{C}(x), \mathcal{P}^k)$ is calculated by combining the Euclidean distance d_e and dot product d_d:

$$\begin{aligned} d_e(\mathcal{C}(x), \mathcal{P}^k) &= \frac{1}{m} \cdot \|\mathcal{C}(x) - \mathcal{P}^k\|_2^2, \\ d_d(\mathcal{C}(x), \mathcal{P}^k) &= \mathcal{C}(x) \cdot \mathcal{P}^k, \\ d\ (\mathcal{C}(x), \mathcal{P}^k) &= d_e(\mathcal{C}(x), p_i^k) - d_d(\mathcal{C}(x), \mathcal{P}^k). \end{aligned} \tag{2}$$

The learning of θ is achieved by minimizing the reciprocal points classification loss based the negative log-probability of the true class k:

$$\mathcal{L}_c(x; \theta, \mathcal{P}) = -\ log\ p(y = k|x, \mathcal{C}, \mathcal{P}). \tag{3}$$

Through minimizing Eq. (3), the reciprocal points classification loss reduces the empirical classification risk through the reciprocal points.[3]

Adversarial Margin Constraint Now we get representation of outlier space of category k. To separate it with known samples in class k as much as possible. We need some restrictions so that the open space can be estimated. Adversarial margin is used to solve this problem.
Considering that space of category k and its outlier space are complementary to each other, the open space risk can be bounded indirectly by constraining the distance between the samples from class k and the reciprocal points $\mathcal{P}^k$ to be smaller than R as follows:

$$\mathcal{L}_o(x; \theta, \mathcal{P}^k, R^k) = max(d_e(\mathcal{C}(x), \mathcal{P}^k) - R, 0), \tag{4}$$

where R is a learnable margin.Through this adversarial mechanism, each known class is pushed to the edge of the finite feature space to the maximum extent, moving each far away from its potential unknown space.[3]

Learning Framework In adversarial reciprocal points learning, the overall loss function combines Eq. (3) and Eq. (4) to handle the empirical classification risk and the open space risk simultaneously:

$$\mathcal{L}(x, y; \theta, \mathcal{P}, R) = \mathcal{L}_c(x; \theta, \mathcal{P}) + \mathcal{L}_o(x; \theta, \mathcal{P}, R), \tag{5}$$

where $\theta, \mathcal{P}, R$ represent the learnable parameters.

3 Experiments

3.1 Dataset and Preprocessing

A good open set recognition classifier should successfully classify the known samples and reject unknown samples in the open world. We choose a widely used academic dataset MNIST, and considers 6 classes as known class and the remaining 4 classes as unknown to evaluate our model.

To better test robustness of ARPL,Cifar 10 and Cifar 100 are used. In preprocessing, We split the dataset into the known and unknown class. The training set only contains samples in the known class, and the testing set contains samples both from the known and unknown classes. Before training the model, we normalized the data image according to their mean value and standard deviation.

3.2 Evaluation Metrics

Simialr to most OSR papers, A threshold-independent metric, the Area Under the Receiver Operating Characteristic (AUROC) curve is used to evaluate accuracy when detecting unknown classes and Open Set Classification Rate (OSCR) is used to evaluate accuracy when when classifying known classes.

3.3 Results

For comparison, we use results on MNIST dataset as a measurment index. As shown in Table 1, the ARPL framework shows a significant boost on the ability to separate unknown classes form known categories. Also, from Table 2, we see that the the algorithm does not sacrifice classification accuracy, it achieves quite satisfying results on both subtasks of OSR problems. For robustness test, we further test ARPL Algorithm on CIFAR10 dataset (with 6 classes unknown) and CIFAR100-50 (randomly pick 50 classes with 4 unknown and 46 known), the results met expectation.

Table 1: The AUROC results of on detecting known and unknown samples. Results are averaged among five randomized trials.

Method	MNIST	CIFAR10	CIFAR100-50
Logistic	83.57 $\pm$0.1	$\cdots$	$\cdots$
SVM	93.50 $\pm$0.1	$\cdots$	$\cdots$
NN	89.12 $\pm$0.1	$\cdots$	$\cdots$
CNN	90.70 $\pm$0.2	$\cdots$	$\cdots$
ARPL	99.34 $\pm$ 0.1	89.01 $\pm$ 0.3	91.20 $\pm$ 0.2

Table 2: The open set classification rate (OSCR) curve results of open set recognition. Results are averaged among five randomized trials.

Method	MNIST	CIFAR10	CIFAR100-50
Logistic	79.85 $\pm$0.1	$\cdots$	$\cdots$
SVM	90.80 $\pm$0.1	$\cdots$	$\cdots$
NN	89.58 $\pm$0.1	$\cdots$	$\cdots$
CNN	91.00 $\pm$0.4	$\cdots$	$\cdots$
ARPL	99.18 $\pm$0.1	85.20 $\pm$0.7	89.74 $\pm$0.5

4 Conclusion

In this project, we first modify several algorithms to solve *Open-Set-Recognition* problems, analyze and compare their results. Traditional classification methods such as SVM,Nearest Neighbours, Logistic and CNN did not perform well on detecting unknown classes from known ones as they only focus on limited parts of known class of the whole open space. Then we reproduce the method ARPL proposed in the paper [2] to solve above queations by learning latent representation of unknown and set retrictions to seperate them with known samples. This method shows huge boost compared with previous models and we test its robustness on larger datasets.
OSR is still a quite new aspect in classification, and by now most methods are tested with relative small datasets, like MNIST, CIFAR and TinyImageNet. In the future, much bigger datasets need to be taken into experiments to better evaluate those methods. Also, most existing methods do not have a clear definition of 'semantic class' which OSR should be focus on. Further, a semantic shift benchmark [1] was proposed to better understand and evaluate OSR. Those two aspects may be where future work lies in.

References

[1] Walter J. Scheirer, Anderson de Rezende Rocha, Archana Sapkota, and Terrance E. Boult. Toward open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013.

[2] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Open-set recognition: a good closed-set classifier is all you need? *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021.

[3] Guangyao Chen, Peixi Peng, Xiangqian Wang, and Yonghong Tian. Adversarial reciprocal points learning for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[4] Pedro R. Mendes Júnior, Roberto M. de Souza, Rafael de O. Werneck, Bernardo V. Stein, Daniel V. Pazinato, Waldir R. de Almeida, Otávio A. B. Penatti, Ricardo da S. Torres, and Anderson Rocha. Nearest neighbors distance ratio open-set classifier. *Machine Learning*, 2017.

[5] T. F. Wu, C. J. Lin, and R. C. Weng. Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research*, 2004.