Failure Detection Based on Anomaly Detection and Multiple-Layer Perceptron Facing Unbalanced Sample Set

Ci Song
Northeastern University at Qinhuangdao Branch, Qinhuangdao 066004, China.

Abstract: As computational performance continues to improve, machine learning is increasingly being used in a variety of areas. Classification problems are one of the most common problems people encounter in their daily lives. However, many classification tasks are confronted with the problem of sample imbalance, which is considered tricky. Although researchers have developed many algorithms for this, problems, such as overfitting, still result in poor classification results in many cases. This paper tries to solve a binary classification with unbalanced sample set applying an idea of combining ready-made anomaly detection and deep learning methods, where anomaly detection algorithms are taken as filters to exclude the effect of samples that are easy to be recognized as the ones from the major category on the final classification done by neural network. This idea is proved more useful on the machine failure detection than using anomaly detection or MLP classifier alone and is believed to be able to serve as a backup or pretest choice in some classification tasks with sample imbalance.

Keywords: Machine Learning; MLP; Sample Imbalance

1. Introduction

In binary classifications, the performance of the classification is mainly measured by Accuracy, which can be unwise facing datasets with unbalanced samples. In practice, classification tasks with unbalanced samples like fraud detection, diagnosis of medical diseases are rather common [1]. To solve such problem, Kubat et al. purposed a solution called one-sided selection, which decrease the noise and redundancy in the large category [2]. Chawla et al. suggest a method of artificially adding samples to the small category to achieve a relatively better balance called SMOTE [3].

This paper, on the other hand, tries to solve such problem using a very simple technique by regarding it as anomaly detection. By comparing different machine learning algorithms like Outlier Analysis based on Probability and Statistics, Isolation Forest and LOF, it attempts to find the method that suits the dataset we choose the best. Besides anomaly detection, this paper investigates the ability of Multiple-layer Perceptron to recognize categories with sparse sample sizes. After that, a strategy that combines of anomaly detection and MLP which uses two layers of anomaly detection as filters and MLP as the final classifier is purposed.

The motivation of this paper is that, anomaly detection distinguishes abnormal points from the normal points, which is pretty similar to what needs to be done in binary classification with unbalanced sample proportion.

2. Data description

This paper applies *Machine operation monitoring dataset* given in *the 2nd Chinese University Big Data Challenge 2022*. The dataset contains 9000 rows of data, among which there are 9697 samples labeled 1 and 303 samples labeled 0, where 1 means that the machine works normally and 0 means otherwise. This paper denotes the samples labeled 0 as positive and samples labeled 1 as negative. As is shown in figure 1, each sample has 8 features, which are ID numbers, specification code, quality level, room temperature, machine temperature, rotate speed, torque and working hours. Using Pearson Correlation Coefficient, room temperature and rotate speed are dropped out of the feature list.

ID number	specification cod	quality level	room tempe	machine tem	rotate speed（r	torque（Nm)	working hours（min)	failure or not
84	L48027	L	296.4	307.4	2833	5.6	213	1
6986	M15740	M	295.8	306.3	1235	76.2	89	1
8047	L48083	L	295.7	306.2	2270	14.6	149	1
4425	M15847	M	296.3	307.1	1534	33.8	151	0
4519	H30402	H	296.3	307.1	1774	25.9	154	0
1877	M15849	M	296.2	307	2119	18.3	159	0
6034	L48170	L	296.2	307	1414	48.3	162	0
5390	M15851	M	296.1	307	1523	42	164	0
7455	L48172	L	296.1	307.1	1651	35.7	167	0
5536	M15853	M	296.1	307.1	1485	36	169	0
7293	M15854	M	296.2	307.2	1168	63.4	172	0
6507	L48175	L	296.3	307.3	1566	35.8	175	0

Figure 1: Raw data

3. Anomaly detection

3.1 Outlier Analysis based on Probability and Statistics

Outlier Analysis based on Probability and Statistics detect the abnormal sample by calculating the mean and variance of each feature and generating the multiple variables Gaussian Probability density function of the form, taking 2-dimensinal feature as an example:

$$f(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)}{\sigma_x^2}+\frac{(y-\mu_y)}{\sigma_y^2}-\frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right)$$

Where ρ is the correlation coefficient. Using probability density function, the points with probability lower than the threshold that is set beforehand shall be detected as abnormal.

After several experiments, it is found that when the contamination is set 0.1, the model achieves the best classification performance, as is shown in the confusion matrix below in figure 3:

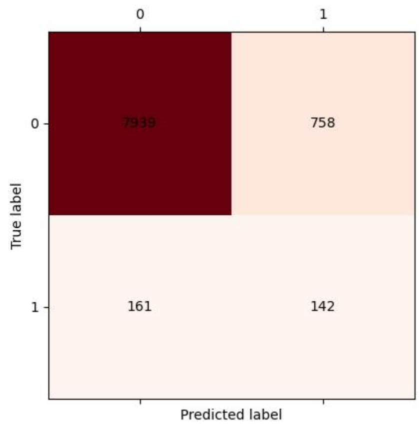

Figure 3: Confusion matrix achieved by Outlier Analysis

based on Probability and Statistics

3.2 Isolation Forest

Isolation Forest defines the points that are more likely to be separated as anomalous data, which are sparsely distributed and far from densely populated groups. [4]. This paper again set the contamination as 0.1, the confusion matrix achieved by iForest is shown in figure 4 below:

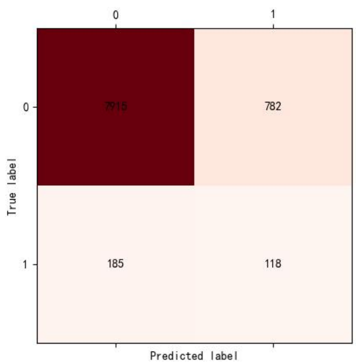

Figure 4: Confusion matrix achieved by Isolation Forest

As is shown in the confusion matrix above, the binary classification task with imbalance in samples performed by Isolation Forest is even worse than what is achieved by Outlier Analysis based on Probability and Statistics.

3.3 LOF algorithm

Local Outlier Factor (LOF) is simple, intuitive, does not require knowing the distribution of the data set, and quantifies the degree of anomalies at each sample point.

After some experiment, parameter n_neibgbors is set as 60 and contamination is still 0.1. Figure 5 is the confusion matrix achieved by LOF, which shows that the classification performance achieved by LOF is the best among three algorithms, still not ideal enough though.

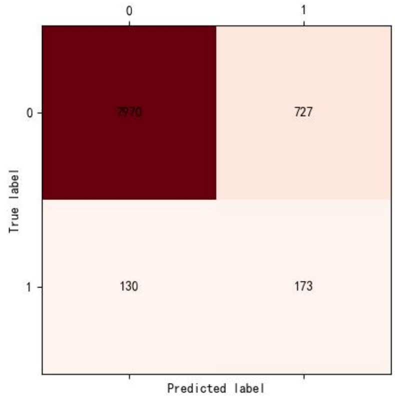

Figure 5: Confusion matrix achieved by LOF

4. Multiple-layer perception

Now, a neural network that can dig out complex relationships is considered. The model used here is Multiple-layer perception, each hidden layer uses “relu” as activation function. The activation function of output layer activation function is "SoftMax", which is meant for classification tasks. The output result is the probability of whether there is a failure or not. The loss function is "binary_crossentropy", the optimizer is "Adam" with a learning rate of 0.00005, and the batch size is set to 256. The structure of the MLP is presented in figure 6:

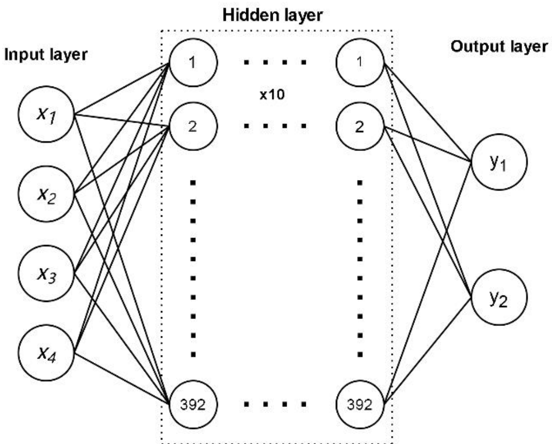

Figure 6: Structure of MLP

It costs only 2 epochs to converge and the result is shown in the confusion matrix in figure 7. Obviously, MLP categorizes all the points into “Normal”, which is not surprising at all. As is mention at the beginning of this paper, the classifier can easily reach high accuracy just by categorizing all the sample into the largest category, which explains the reason why MLP converges so fast and achieves such poor performance.

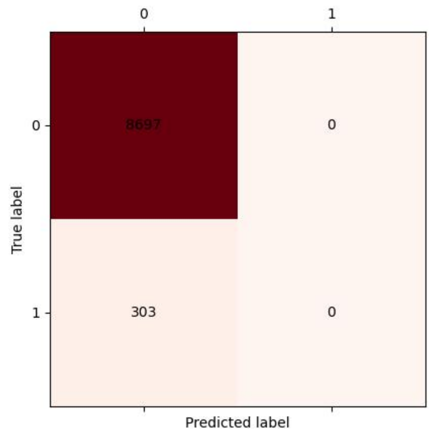

Figure 7: Confusion matrix achieved by MLP

5. Combination of anomaly detection and MLP

Since the direct cause of MLP classifying all the samples as "Normal" is the large difference in positive and negative sample sizes, this section purposes a strategy that first uses anomaly detection to sift out samples that are obviously "Normal" and use MLP to binarily classify the rest, which achieves better performance than using any of the four methods alone.

5.1 Filter

This paper uses two times of anomaly detection as filters, each filter sets the contamination as 0.5, by which the most "Abnormal points" can be preserved each time. To choose the algorithm of each filter, this paper tries different combinations of LOF and Outlier Analysis based on Probability and Statistics, the results are shown in the table 1. L is denoted as LOF and P is denoted as Outlier Analysis based on Probability and Statistics.

Table 1: Comparison between combinations of L and P

Combination	TP	FN	FP	TN
L+L	6634	2063	116	187
L+P	6644	2053	106	197
P+L	6648	2049	102	201
P+P	6630	2067	120	183

The table shows an obvious advantage of using Outlier Analysis based on Probability and Statistics as the first filter and LOF as the second over other combinations.

5.2 Final classifier

After the samples have gone through two filters chosen before, it is believed that MLP can exert its ability to approximate complex relationships.

The MLP built in this experiment has 4 hidden layers. the model is trained for 1000 epochs. The change of loss and accuracy are visualized in figure 7:

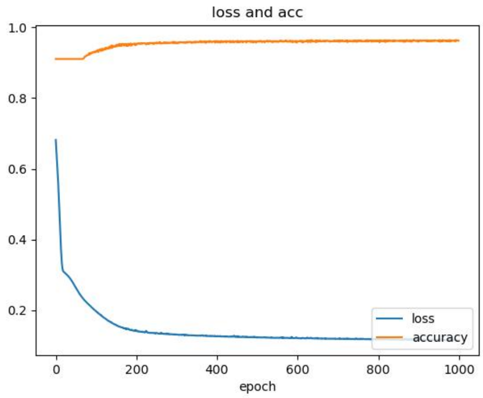

Figure 7: Change of loss and accuracy during the training

From the chart, it can be seen that the accuracy does not change at all at the beginning. It is obviously because the MLP again classifies all the samples into “Normal”. However, instead of staying still like what happens with using MLP directly, the accuracy rises after a few epochs, which indicates that our strategy of using filters is effective. The confusion matrix achieved by our combination model is shown in figure 8.

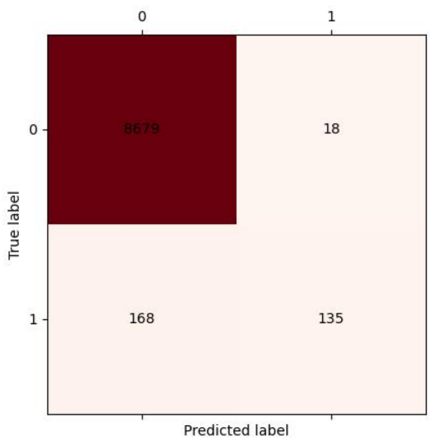

Figure 8: Confusion matrix achieved by combination model

Although the result is not perfect and more than half of the “Abnormal” points is classified into “Normal”, the “Abnormal” points recognized by combination model much more precise.

Table 2 is made to compare the combination model with using anomaly detection directly:

Table 2: TP, TN, precision, recall and accuracy

model	TP	TN	Precision	Recall	Accuracy
Probabilistic and statistical model	7939	142	0.158	0.469	0.898
Isolation Forest	7889	92	0.131	0.389	0.893
LOF	7970	173	0.192	0.571	0.905
Combination model	8679	135	0.882	0.446	0.979

Combination of Anomaly detection and MLP makes a little sacrifice of “recall” in exchange for high precision and accuracy. In general, if it is not a case where it is better to kill a thousand mistakes than miss one, the effect achieved by the combination model is significantly better than anomaly detection used alone.

6. Conclusion

This paper puts forward a strategy for binary classification task with unbalanced samples. By first using properly chosen anomaly detection methods as two or several layers of filters to sift out samples that are easy to recognized as the majority, the unbalanced sample may become more balanced for Neural Network to better approximate the mapping functions between features and labels. This strategy is extremely simple to carry out and proved effective on the *Machine operation monitoring data set.* It works especially well with reducing the mistakes of classifying the samples of major category into the minority.

References

[1] Li DC, Liu CW, Hu SC. A learning method for the class imbalance problem with medical data sets. Comput Biol Med. 2010 May;40(5):509-18.

[2] Kubat M, Matwin S. Addressing the curse of imbalanced training sets: one-sided selection[C]// Icml. 1997, 97(1): 179.

[3] Chawla NV, Bowyer KW, Lawrence OH, et al. SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 2002; 16: 321—357.

[4] Liu, FT, et al. Isolation Forest[J]. IEEE Data Mining, 2008.

[5] Faris H, Aljarah I, Mirjalili S. Training feedforward neural networks using multi-verse optimizer for binary classification problems[J]. Applied Intelligence, 2016, 45(2): 322-332.

[6] Wang Z, Zhang M, Wang D, et al. Failure prediction using machine learning and time series in optical network[J]. Optics Express, 2017, 25(16): 18553-18565.