FREE LUNCH FOR FEW-SHOT LEARNING: DISTRIBUTION CALIBRATION REPRODUCABILITY CHALLENGE

Yijun Chen
ID 31789501
yc16g20@soton.ac.uk

Shuning Ling
ID 31543405
sl4m20@soton.ac.uk

Pasinpat Vitoochuleechoti
ID 31906257
pv1u20@soton.ac.uk

ABSTRACT

FREE LUNCH FOR FEW-SHOT LEARNING: Distribution Calibration written by Shuo Yang, Lu Liu, Min Xu is to transfer statistics from base classes with enough examples to calibrate the distribution of these few-sample classes, and then to draw a sufficient number of examples from the calibrated distribution to expand the input of the classifier. The calibrated distribution is then drawn from a sufficient number of examples to expand the input to the classifier Yang et al. (2021). By running the Distribution Calibration code in the appendix of this paper and pre-training the data, we will confirm whether the results mitigate the overfitting phenomenon in few-sample learning, as claimed in this paper. By calculating the accuracy of SVM and logistic regression, Tukey transformation and the presence or absence of generated features, we see that Distribution Calibration does have some improvement on the overfitting problem.

1 INTRODUCTION

Due to the high cost of extracting models from large amounts of data and conducting research, researchers have been trying to train extracted models with small amounts of data (Finn et al., 2017; Snell et al., 2017). However, when the amount of data is too small, it tends to cause overfitting as the model tends to minimise traning loss on these few samples. For these reasons, and subject to the two conditions of using a small amount of data and avoiding overfitting, this experiment considers a biased distribution calibration method that will be used to improve the problem. Instead of calibrating the spatial distribution of the original data, we attempt to calibrate the distribution in feature space, which is much easier to calibrate because of its lower dimensionality (Xian et al., 2018). We also assumed that the distribution of the data in each each dimension conformed to a Gaussian distribution and observed that some similar classes had similar means and variances, so that the means and variances of the Gaussian distribution could be transferred between similar classes. Based on these observations, we calibrated the distribution with fewer sample classes by transferring statistics from base classes with sufficient examples, and then estimated the corrected samples.

2 MAIN APPROACH

2.1 CALIBRATING STATISTICS OF THE NOVEL CLASSES

2.1.1 TUKEY'S LADDER OF POWERS TRANSFORMATION

Tukey' ladder of power transformation can make the distribution of features more like a Gaussian distribution. Its transformation is a family of power transformations that reduce the skewness of the distribution and make it more Gaussian-like.

2.1.2 CALIBRATION THROUGH STAISTIC TRANSFER

We assume that the distribution of the characteristics of the base class is consistent with a Gaussian distribution. We transfer the statistics of the base class to the new class, and these statistics are then estimated more accurately on sufficient data. The transfer is based on the Euclidean distance

Table 1: Result of Logistic Regression and SVM with Distribution Calibration and 3 selected best method

Method	miniImageNet		CUB	
	5way1shot	5way5shot	5way1shot	5way5shot
Meta-SGD (Yang et al., 2021)	50.47±1.87	64.03±0.94	53.34±0.97	76.59±0.82
Negative-Cosine (Yang et al., 2021)	62.33±0.82	80.94±0.59	72.66±0.85	89.40±0.43
TriNet (Yang et al., 2021)	58.12±1.37	76.92±0.69	69.61±0.46	84.10±0.35
SVM with DC	68.1205	83.2844	80.4075	90.8077
Logistic Regression with DC	68.1752	83.5656	80.2893	90.8107

between the feature space of the new class and the mean of the features of the base class. Specifically, we wish to select the top k base classes that are closer to the feature class (Yang et al., 2021).

3 EXPERIMENTS

On this experiments, we attempt to re-implement experiments from Yang et al. (2021) using their Distribution Calibration and Tukey's Ladder of Powers Transformation. Section 3.1 discusses experimental setup on CUB and miniImageNet datasets and implementation details. Section 3.2 discusses on reproduce experiment and result compare with other state-of-the-art machine learning method. On Section 3.3 and 3.4 will discusses on effects of feature transformation and α hyper-parameter, respectively. Moreover, section 3.5 will discuss on further result from our more experiments base on Yang et al. (2021) implementation. Our code is available at `https://github.com/COMP6248-Reproducability-Challenge/Free-Lunch-For-Few-Shot-Learning`

3.1 EXPERIMENTAL SETUP

3.1.1 DATASETS AND EVALUATION METRIC

The CUB datasets contain 200 different classes of birds and miniImageNet datasets contain 100 diverse classes use for evaluate on experiments based on code produced by Yang et al. (2021). These both datasets classes come with a total of 11,788 image of size $84 \times 84 \times 3$ and 600 samples per class of image of size $84 \times 84 \times 3$, respectively. In addition, Yang et al. (2021) demonstrated they work use the top-1 to measure the performance of method on the evaluation metric, and the average classification accuracy is the result from evaluating over 10,000 task that we use same evaluation metric as Yang et al. (2021). Yang et al. (2021) code available at `https://github.com/ShuoYang-1998/Few_Shot_Distribution_Calibration`

3.1.2 IMPLEMENTATION DETAILS

On this experiment, we use Distribution Calibration and Tukey's Ladder of Powers Transformation based on Yang et al. (2021) code and use same hyper-parameter for all datasets that λ = 0.5, the number of generated features is 750, k = 2, α is 0.21 and 0.3 for miniImageNet and CUB, respectively. In section 3.2, we evaluate by using Logistic Regression and SVM library from scikit-learn with the default settings. On Logistic Regression experimentation, we change values of power λ in Turkey transformation, disable it and disable generated features to evaluate in section 3.3. In addition, we experiment by change α Hyper-parameter in Distribution Calibration from default setting by use Logistic Regression in section 3.4. We tried Gaussian Naive Bayes from scikit-learn with default settings and use Distribution Calibration and Tukey's Ladder of Powers Transformation on section 3.5. In section 3.5 we evaluated Baseline method and Baseline with Distribution Calibration by using L2 Normalize and not use on different λ. Last, all experiments take time to run every 5way1shot method 5 to 12 hour and take time to run every 5way5shot method 19 to 40 hours (using AMD Ryzen 7 4800HS 2.90 GHz).

Table 2: Ablation study on miniImageNet 5way1shot and 5way5shot showing accuracy (%)

Turkey transformation	Training with generated features	miniImageNet	
		5way1shot	5way5shot
✗	✗	60.0607	81.4237
✓	✗	64.6565	83.4289
✗	✓	64.1015	81.1565
✓	✓	68.1752	83.5656

3.2 Comparision to State-of-the-art

The experiment in this section selects 3 best method from the method that Yang et al. (2021) compare to Logistic Regression and SVM with Distribution Calibration (DC) to compare with our Logistic Regression and SVM with Distribution Calibration implementation as the result illustrated in Table 1. In Table 1 shows percent of 5way1shot and 5way5shot classification accuracy on miniImageNet and CUB with 95% confidence intervals except SVM with DC and Logistic Regression with DC method. We can see simple machine learning classifier with DC like Logistic Regression and SVM have accuracy outperform all other machine learning.

3.3 Effects of feature transformation

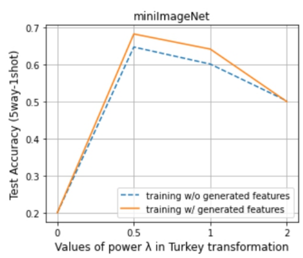

Figure 1: Values of power λ in Turkey transformation

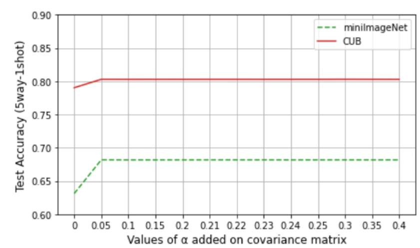

Figure 2: The effect of different values of α

In this section, we evaluate Logistic Regression with DC by have disable and enable Tukey's Ladder of Powers Transformation and disable and enable generated features as present in Table 2. Another experiment, we observe on values of power λ in Tukey transformation changing with and without generated features as illustrated in Figure 1. On Table 2, we can see when Tukey's Ladder of Powers Transformation and generated features enable accuracy will increase. Figure 1 shows the accuracy with increase follow λ until λ is 0.5. Then, the accuracy will decrease (The code will error when experiment on $-\lambda$).

3.4 α Hyper-parameter

Figure 2 presents the plots of different value of α implement by Logistic Regression with DC and default setting and hyper-parameter. We will see in Figure 2, the accuracy of the unique value of α on this implement is rising until 0.05 then it stable different from Yang et al. (2021) that accuracy will decrease since α is 0.25.

3.5 Further Experiments

We select other simple machine learning that is gaussian naive bayes to implement with Distribution Calibration and Tukey's Ladder of Powers Transformation to find different from Yang et al. (2021). Table 3 presents the result of Gaussian Naive Bayes with DC and Tukey Transformation implement by 5way1shot and 5way5shot on miniImageNet and CUB datasets. Its accuracy has opposite values to other simple machine learning with DC and Tukey Transformation. Moreover, we use L2 Normalize with Baseline method with and without DC on different λ to observe the effect of L2

Table 3: 5way1shot and 5way5shot classification accuracy (%) on miniImageNet and CUB by Gaussian Naive Bayes with DC

Method	miniImageNet		CUB	
	5way1shot	5way5shot	5way1shot	5way5shot
Gaussian Naive Bayes with DC	30.3648	30.5739	38.9013	39.6248

Table 4: 5way1shot classification accuracy (%) on miniImageNet by Baseline and Baseline with DC with L2 normalize

Method	λ	L2 Normalize	5way1shot
Baseline with DC	0.5	w/o	68.2073
Baseline with DC	0.5	w	67.3253
Baseline	0.5	w/o	64.6253
Baseline	1	w/o	59.0667
Baseline	0.5	w	65.1833
Baseline	1	w	66.3907

Normalize as the result illustrated on Table 4. We will see the result from Table 4, the L2 Normalize did not much impact to the accuracy and it decrease a little of accuracy when compare the method that not using L2 Normalize and Baseline with DC is producing the best accuracy.

4 CONCLUSION

In this project it is first possible to reproduce the results about SVM and Logistic Regression with Distribution Calibration. All the code and pre-processed data given in the original paper were reproduced, including both the miniImageNet and CUB data sets. Most of the results obtained are within the range of those mentioned in the paper being reproduced, indicating that indeed Distribution Calibration is well optimised for logistic regression problems, as stated by the authors. Then, the code for the Tukey transformation of λ from 0 to 2 was reproduced and the results obtained were indeed more accurate after Distribution Calibration. It is worth noting that Distribution Calibration is very costly in terms of time, as the Tukey transformation without generated features may only take a dozen minutes to run through λ = 0.5 way = 5, shot = 1 parameter, but with generated features this value increases to several hours. Next, a comparison of the effect of the hyperparameters on the accuracy of the optimization shows that after the hyperparameter α reaches 0.05, the effect on the accuracy of the optimization becomes very small, so a fixed hyperparameter result of 0.05 is sufficient. Finally, we tested Distribution Calibration on Gaussian Naive Bayes and the results were not very good. Due to time and hardware constraints, the effect of Backbones optimisation was not tested. Meanwhile, a comparison was made between Distribution Calibration and L2 Normalize and it was found that Distribution Calibration (almost 4% improvement) was better than L2 (about 1% improvement). Distribution Calibration alone is better than Distribution Calibration plus L2 Normalize (about 3% improvement).In summary, Distribution Calibration is an excellent method of optimisation.

REFERENCES

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks, 2017.

Jake Snell, Kevin Swersky, and Richard S. Zemel. Prototypical networks for few-shot learning, 2017.

Yongqin Xian, Tobias Lorenz, Bernt Schiele, and Zeynep Akata. Feature generating networks for zero-shot learning, 2018.

Shuo Yang, Lu Liu, and Min Xu. Free lunch for few-shot learning: Distribution calibration, 2021.