A Study of the Ratio of Labeled and Unlabeled Data of Semi-supervised Learning with Ladder Networks

Xiao Ji, Zining Fan, Zhuoyue Xing

Abstract

In this project, we would like to explore the ratio of labeled and unlabeled data in semi-supervised learning using Ladder Network model. We used MNIST dataset and put aside some images as test data at the beginning. We then used 1000 images as the labeled data, and changed the amount the unlabeled data. The test accuracy reached the highest when the labeled/unlabeled ratio is 1:90.

1 Introduction

1.1 Related Work

Deep Learning has shown remarkable success in image recognition[1]. We could achieve image classification with deep convolutional neural networks[2]. However, these models need to be trained with supervised learning which requires a large amount of labeled data[3]. By showing the models only labeled images, we limit ourselves from making use of unlabeled images. To avoid this, we could combine labeled and unlabeled data and fit a semi-supervised learning model.

The model we will use in this study is Ladder Networks[4]. Ladder network was proposed by Valpola in 2015 for unsupervised learning. Antti then extended it by combining the model with supervision and reaches state-of-art performance in semi-supervised MNIST anf CIFAR-10 classification in 2015. We would like to learn how Antti constructed the ladder network algorithm and the impact of different parameter settings on the test accuracy.

1.2 Problem Formulation

The goal of our paper is to study two main problems. First, we will change the ratio of the amount of labeled and unlabelled data in the training set and find the optimal ratio which maximizes test accuracy. Second, we would like to investigate the impact of the Gaussian noise of the ladder network encoder on the test accuracy. More explanations of the model is in 2.2.
In summary, we will conduct the study by fitting ladder network models in the following ways:

1. Fixed Gaussian noise with different ratio of labelled data to unlabeled data;
2. No noise with different ratio of labelled data to unlabeled data;
3. Different Gaussian noise with the fixed optimal ratio of labeled data to unlabeled data.

2 Dataset to be used and method proposed

2.1 MNIST Dataset

The study will be conducted using the MNIST Database of handwritten digit, which has a training set of 60,000 examples, and a test set of 10,000 examples.

In our project, we would like to test the ratio of the labelled and unlabelled data. So, we put aside the 10,000 test samples at first; then we will select some examples from the training and keep their labels, and ignore the labels of the rest of the training set. We will use these labeled images, together with different number of unlabeled images to train our model, and predict the labels of the test images. Then We will add the unlabeled data and gradually increase the ratio of labeled and unlabeled data.

2.2 Ladder Network

The model we used in our project is ladder network[4]. The basic structure of ladder network is shown as follows:

1. A feedforward model as encoder which serves supervised learning
2. A decoder which could invert the mapping of each layer serving unsupervised learning
3. Train the whole Ladder network by minimizing the sum of all the cost function terms

Algorithm 1 Ladder Network

Require: $x(n)$
Corrupted encoder and classifier
$\tilde{\mathbf{h}}^{(0)} \leftarrow \tilde{\mathbf{z}}^{(0)} \leftarrow \mathbf{x}(n) + noise$
for $l = 1$ to L **do**
    $\tilde{\mathbf{z}}^{(l)} \leftarrow \text{batchnorm}\left(\mathbf{W}^{(l)}\tilde{\mathbf{h}}^{(l-1)}\right) + \text{noise}$
    $\tilde{\mathbf{h}}^{(l)} \leftarrow \text{activation}\left(\gamma^{(l)} \odot \left(\tilde{\mathbf{z}}^{(l)} + \beta^{(l)}\right)\right)$
end for
$P(\tilde{\mathbf{y}}|\mathbf{x}) \leftarrow \tilde{\mathbf{h}}^{(L)}$
Clean encoder (for denoising targets)
$\mathbf{h}^{(0)} \leftarrow \mathbf{z}^{(0)} \leftarrow \mathbf{x}(n)$
for $l = 1$ to L **do**
    $\mathbf{z}_{\text{pre}} \leftarrow \mathbf{W}^{(l)}\mathbf{h}^{(l-1)}$
    $\boldsymbol{\mu}^{(l)} \leftarrow \text{ batchmean }\left(\mathbf{z}_{\text{pre}}^{(l)}\right)$
    $\boldsymbol{\sigma}^{(l)} \leftarrow \text{ bat chstd }\left(\mathbf{z}_{\text{pre}}^{(l)}\right)$
    $\mathbf{z}^{(l)} \leftarrow \text{ batchnorm }\left(\mathbf{z}_{\text{pre}}^{(l)}\right.$
    $\mathbf{h}^{(l)} \leftarrow \text{ activation }\left(\boldsymbol{\gamma}^{(l)} \odot \left(\mathbf{z}^{(l)} + \boldsymbol{\beta}^{(l)}\right)\right)$
end for
Decoder and denoising
for $1 = \text{L to } 0$ **do**
    if $1 = \text{L}$ **then**
        $\mathbf{u}^{(L)} \leftarrow \text{ batchnorm }\left(\tilde{\mathbf{h}}^{(L)}\right)$
    else
        $\mathbf{u}^{(l)} \leftarrow \text{batchnorm}\left(\mathbf{V}^{(l+1)}\hat{\mathbf{z}}^{(l+1)}\right)$
    end if
    $\forall i : \hat{z}_i^{(l)} \leftarrow g\left(\tilde{z}_i^{(l)}, u_i^{(l)}\right)$
    $\forall i : \hat{z}_{i,\text{BN}}^{(l)} \leftarrow \frac{\hat{z}_i^{(l)} - \mu_i^{(l)}}{\sigma_i^{(l)}}$
end for
Cost function C for training
$\text{C} \leftarrow 0$
if $t(n)$ **then**
    $\text{C} \leftarrow -\log P(\tilde{\mathbf{y}} = t(n)|\mathbf{x}(n))$
end if
$\text{C} \leftarrow \text{C} + \sum_{l=0}^{L} \lambda_l \left\|\mathbf{z}^{(l)} - \hat{\mathbf{z}}_{\text{BN}}^{(l)}\right\|^2$ =0

2.3 Change the Ratio

After building the ladder network model, we will use 1000 labeled data and train our model repeatedly by changing the ratio of labeled data to unlabeled data from 1:1 to 1:500. We want to find the optimal ratio that maximize test accuracy. The standard deviation of the Gaussian noise is set to be 0.3.

2.4 Change the Gaussian noise

After analyzing the ratio of labeled and unlabeled data, we will test the impact of the Gaussian noise in the encoder. We will first remove the noise and see how the test accuracy will change with different labeled/unlabeled ratios. Then, we will fix the ratio at the optimal level and change the standard deviation of the noise. Again, we want to find the optimal standard deviation which gives the best test prediction.

3 Results

We used the MNIST dataset from keras.datasets package to train and test the ladder network model. The code for all the experiments is available at https://github.com/ZiningFan00/semi-supervised-learning.

3.1 Gaussian noise with different ratio of labelled data to unlabeled data

To optimize the radio of unlabeled data to all training data in our model, we randomly selected 1000 images as labeled data and changed the size of unlabeled data. The sizes of unlabeled data we used in our model were 1000, 2000, 5000, 10000, 20000 and 50000. We trained all models for 3 epochs. Figure 1 shows the test accuracy of different models.

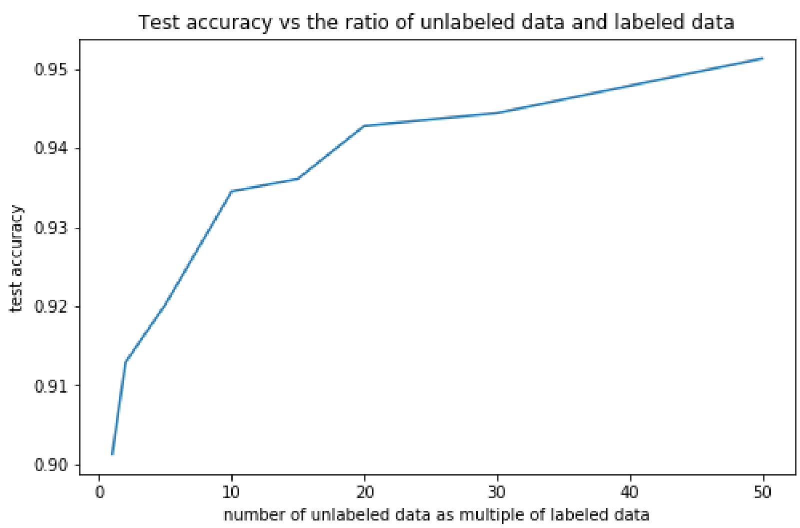

Figure 1: number of labeled images = 1000

In Figure 1, we can see that the test accuracy reaches 90% even when we used only 1000 labeled images and 1000 unlabeled images. Also, the test accuracy keeps increasing as the size of unlabeled dataset goes up. Thus, we decided to use only 500 labeled data in order to explore a larger range of the labeled and unlabeled data ratio.

Figure 2 shows the test accuracy of using the combination of 500 labeled and 1 to 59500 unlabeled data, which provides the range of labeled/unlabeled ratio from 500:1 to 1:119. We could not set the

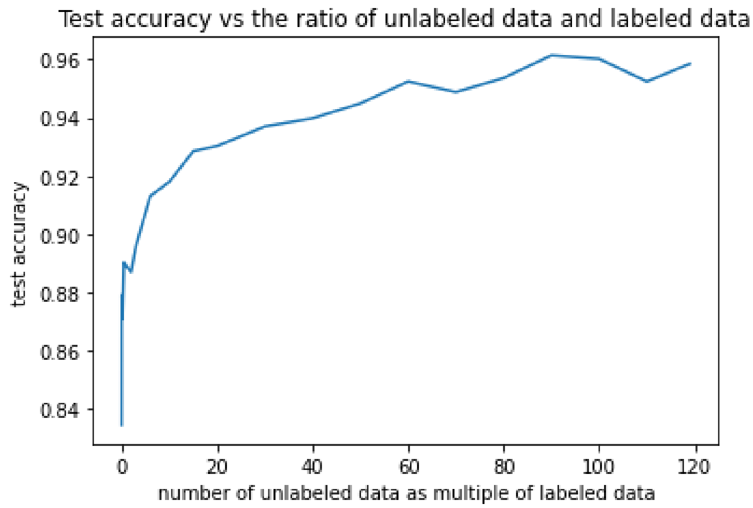

Figure 2: number of labeled images = 500

number of unlabeled data to 0 due to the setting of LadderNet model. We ran 10 epoches for each model.

The starting test accuracy trained by 500 labeled data and 1 unlabeled data is 74%. However, the test accuracy increased to 81% when we used 2 unlabeled data. It is impossible that adding one more image will improve the test accuracy by 7%. This is a limitation of the package and we should start from 2 unlabeled images. The test accuracy increased rapidly to 93% when labeled/unlabeled ratio increases to 1:20. This improvement makes sense because we added an unsupervised learning targets on every layer which captures more relevant details than a supervised learning model on the top layer. As we included more unlabeled data, the test accuracy keeps increasing but at a lower speed. The reason may be that the 20 x 500 additional unlabeled images have already given us many important details, so adding more unlabeled images won't provide much valuable information. The test accuracy hits the highest point when the labeled/unlabeled ratio is 1:90, with a value of 96%.

The next question is: is the optimal ratio relevant to the number of labeled images? We then ran the same algorithm using 100 (Figure 3) and 300 labeled images (Figure 4). If we use 100 labeled images and increase the number of unlabeled images, the test accuracy goes from 72% to 91% and the optimal unlabeled/labeled ratio is 85. If we use 300 labeled images, the test accuracy goes from 81% to 95% and the optimal ratio is 130. However, the test accuracy increases slowly after the ratio becomes 80. Thus, the optimal unlabeled/labeled ratio for MNIST dataset using ladder network is around 80 - 90 and seems not very relevant to the number of unlabeled data.

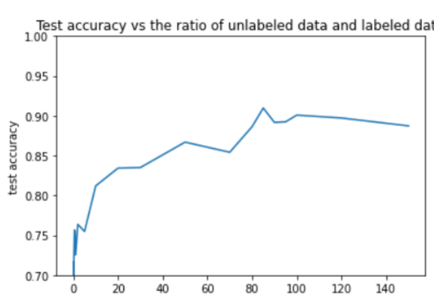

Figure 3: number of labeled images = 100

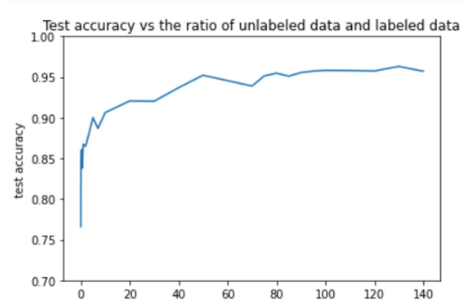

Figure 4: number of labeled images = 300

3.2 No noise with different ratio of labelled data to unlabeled data

To investigate the effect of noise in the ladder network model, we first set the noise to zero and trained the model with the same set of labeled/unlabeled ratios as 3.1. Figure 5 shows the test accuracy. Again, we should ignore the first result (unlabeled/labeled ratio = 500:1). The test accuracy fluctuates around 87%.

As can be seen from the algorithm, the ladder network model adds noise to each layer ($\tilde{z}^{(i)} = z^{(i)} + noise$) and learn a decoder ($\hat{z}^{(i)}$) to minimize the difference between $\hat{z}^{(i)}$ and $z^{(i)}$. This is how ladder network makes use of unlabeled data to learn the structure of input. When we set the noise to zero, $\tilde{z}^{(i)}$ is just $z^{(i)}$ and there is no learning anymore. Thus, the test accuracy won't improve no matter how we change the labeled/unlabeled ratio.

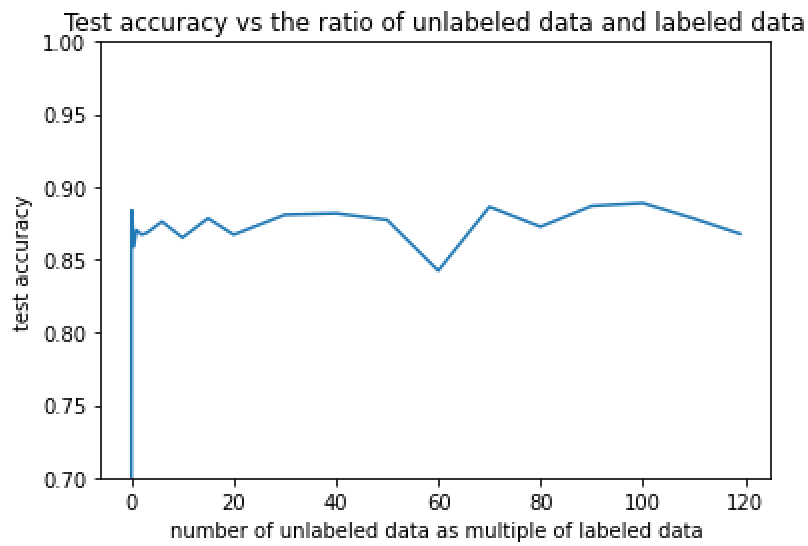

Figure 5:

3.3 Different Gaussian noise with the fixed optimal ratio of labeled data to unlabeled data

We then changed the standard deviation of the Gaussian noise while keeping the labeled/unlabeled ratio to be the optimal value of 1:90. Figure 6 shows the trend of test accuracy as the standard deviation of the noise increases. We got the best test accuracy of 96% when the standard deviation is around 0.3. The optimal standard deviation is related to the data input and would change if we fit a different dataset.

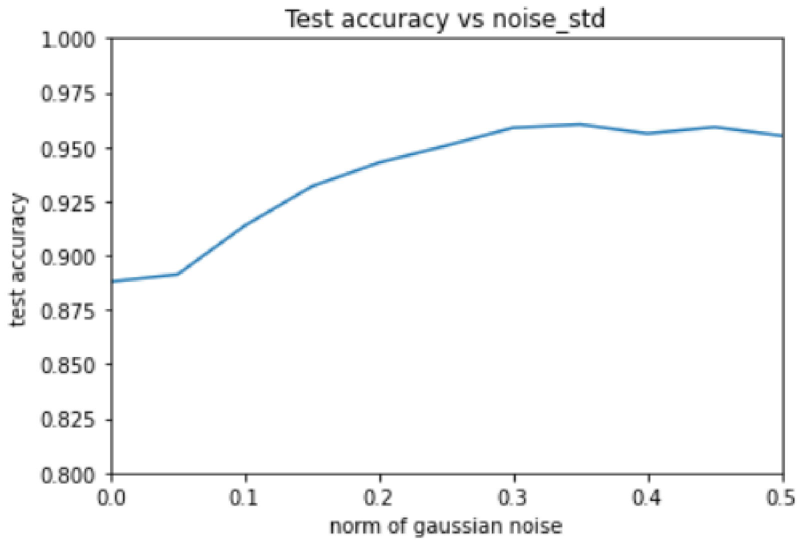

Figure 6:

4 Conclusion and Future Goals

In this project, we deeply analyzed the ladder network model and conducted experiments to study the impact of labeled/unlabeled ratio and noise on test accuracy. We found that adding unlabeled data to the model substantially increases the prediction accuracy and such improvement depends heavily on how we construct the model and set hyperparameters like the labeled/unlabeled ratio and the standard deviation of the noise.

In the future, we can fit the ladder network model on other datasets to see if the optimal labeled/unlabeled ratio keeps the same. Also, we can try different semi-supervised learning models using the same dataset to study how the optimal labeled/unlabeled ratio changes with different models. Last but not least, we can include other noise like dropout, early stop and data augmentations to the ladder network model to see if the prediction accuracy could be further improved.

References

[1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*. 2012, pp. 1097–1105.

[2] Karen Simonyan and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556* (2014).

[3] Mingxing Tan and Quoc V Le. "Efficientnet: Rethinking model scaling for convolutional neural networks". In: *arXiv preprint arXiv:1905.11946* (2019).

[4] Antti Rasmus et al. "Semi-supervised learning with ladder networks". In: *Advances in neural information processing systems*. 2015, pp. 3546–3554.