Project for classifying images from Pinterest

Bin Hwang, Zhiwei Wang, Xinyu Lei

Abstract

The project is to apply deep learning techniques to classify the Pinterest images. Training data set is available from the researchers in College of Education of MSU. These data contains the downloaded educational pins from sampled teachers. We clean the duplicate data, renormalized, rescaled and balanced the dataset into a workable training data and testing data. We apply ANN/DNN and CNN to reached the best 30% accuracy. To better improve classification, based on our assumption that human would first understand concrete materials in figures, we built another classifier to distinguish the images with characters and numbers as to others without . We labeled 1200 images and the classifier reached 80% accuracy.

1 Introduction

In the education domain, lots of educators use the social media site Pinterest to share their teaching skills with colleagues. The information on Pinterest takes the form of images. To analyze the information shared by the educators, the understanding of images is very important.

Considering the complexity of the images, we apply the deep learning methods to understand the them. The images were classified to 9 predefined categories. Our classifiers reached the best of 30% in terms of accuracy.

There are three main challenges that makes the classification task especially difficult. First, the 9 categories are very abstract educational concepts, which is very difficult to classify even for a human being outside the education domain. Second, the amount training data provided is too small to train a deep structure to learn the abstract concept. Third, the training data is very unbalanced as Fig. 1 shows. The contribution of our work are:

- Use the PCA analysis to help find out the best dimensions to classify the images. See Fig. 1

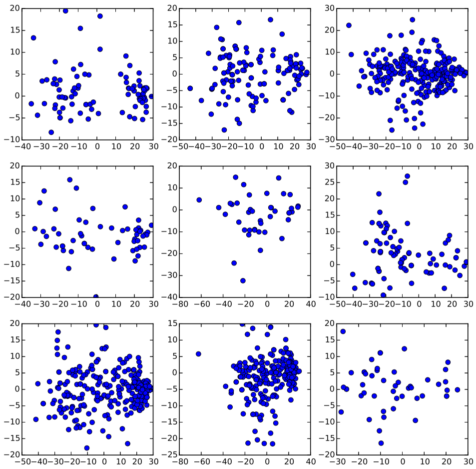

Figure 1: PCA analysis result

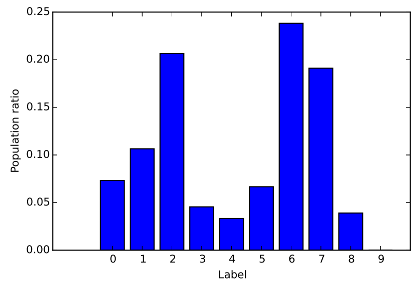

Figure 2: Unbalanced data

- Explore the performance of DNN and CNN on simple binary images classification which might help to understand the abstract concepts.
- Explore the performance of DNN and CNN performance on images classification with

very abstract categories.

The remaining structure of the report is organized as follows: Section 2 provides the formal problem definition and related work. Section 3 provides the two deep learning models'(DNN and CNN) implemented in the project. Section 4 shows the experiments and the results. Section 5 provides discussion of the experiment result and possible future work.

2 Problem Statement

The data consists 1266 images belonging to 9 categories. We will implement deep learning method to build a classifier which can classify a image to one of the 9 categories. This is known as a classification task. Classification is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known.

2.1 Related Work

There are several base classification methods including decision tree classifier [1], nearest-neighbor classifier [2], neural networks [3-4], support vector machines [5], etc. Each classifier owns some advantages in special application scenarios. Since the deep neural network (DNN) shows it superiority in image classification applications recently [6], DNN method is adopted in this paper. Among different neural network models, recurrent neural networks (RNNs) [3] and convolutional neural networks (CNNs) [4] are two popular neural network models. The RNNs model is good at handwriting recognition and speech recognition, whereas the CNNs model is especially suitable for dealing with image classification. In this paper, we investigate both the DNN model and the CNNs model in Pinterest image classification task.

3 Methodology

We use 1266 data points at beginning data set and transfer it into 2268 data points by duplicating small data set for a given label. We split the data into 80% training data and 20% testing data. From Fig. (3) we start with data from Education department and renormalize, reshape and balance data in data preprocessing. We than put data into different methods as ANN/DNN and CNN to build models,

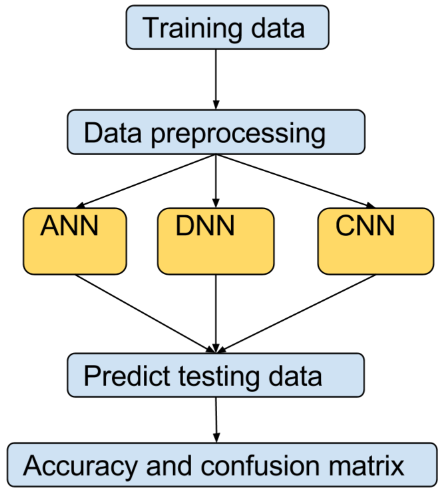

Figure 3: Schematic for data processing and model constructing.

and use these models to predict testing data. After that, we calculate accuracy and confusion matrix from them.

We use square loss to evaluate ANN/DNN and cross entropy loss to evaluate CNN:

$$S_d = (y - f(X))^2 \tag{1}$$

and

$$S_c = -t\ln(f(X)) - (1-t)\ln(1-f(X)), \tag{2}$$

where y is the data label for each data points, X is the training data, $t = (1+y)/2$ and $f(X)$ is the propagator function to final layer. Both of them are with regression loss:

$$R = \lambda(\omega - \omega_0)^2 \tag{3}$$

with ω is the weighting and λ is the regression coefficient.

3.1 ANN/DNN

Artificial neural networking and deep neural networking are learning algorithms that attemt to model high level abstractions in data. Deep learning exploits this idea of hierarchical explanatory factors where higher level, more abstract concepts are learned from the lower level ones. The algorithm can be written as following :

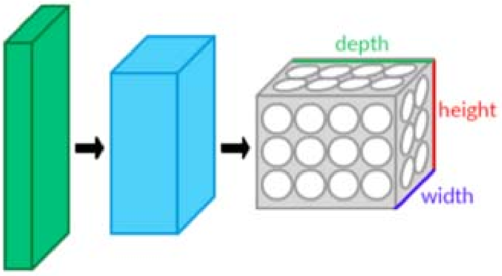

Figure 4: CNN Layers in Graph

1 Initial condition : weight, layer, layer size.

2 forward propagator and calculate loss function.

3 back propagator and update weight.

4 repeat step 2, 3 till loss goal or iteration limit.

3.2 CNN

Shown in Fig. 4 is the CNN layers in graphic representation. Fig. 5 is the corresponding networks structure in Tensorflow. Each node in Fig 5 represents an operation (or computation), the output of the former layer is represented as tensors, which are feed to the next layer. The graph also illustrates why we call the library "Tensorflow". The functions of each layer is clearly illustrated in Table 1. Our CNN model is similar to the model used in the Cifar10 project. One may refer [7] for more detailed mathematical information.

In Fig. (6) shows different regression coefficient λ vs accuracy with three layers. Size for two hidden layers are $[20, 10]$. We find the best accuracy 21.12% is when $\lambda \approx 0.001$ so this coefficient would be used in different layers learning. For large number regression as $\lambda = 0.1$ we can find the model focus on the weight loss than correct prediction.

4 Experimental Evaluation

4.1 ANN/DNN

From Fig. (7) we presents different layers vs accuracy with regression coefficient $\lambda = 0.001$. Size for one hidden layer is $[10]$, two hidden layers are $[20, 10]$, three layers are $[30, 20, 10]$ and four layers are [40,30,20,10]. One can recognize that when we use four layers as hidden layer is equal

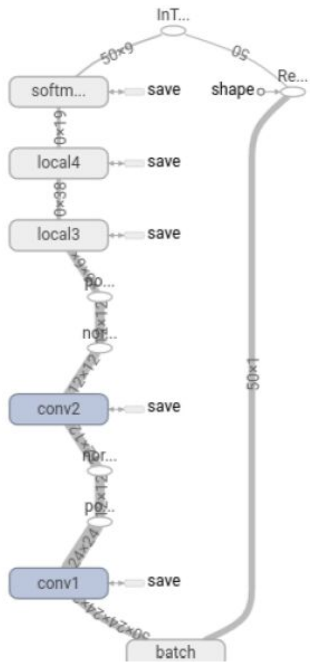

Figure 5: CNN Layers in Tensorflow

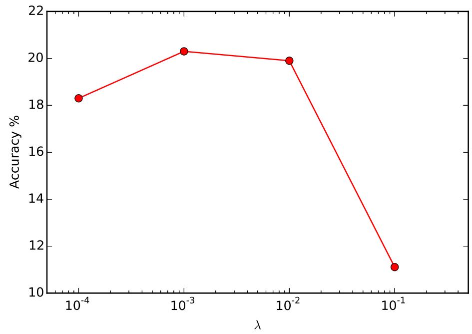

Figure 6: different regression coefficient λ vs accuracy with three layers. Size for two hidden layers are $[20, 10]$.

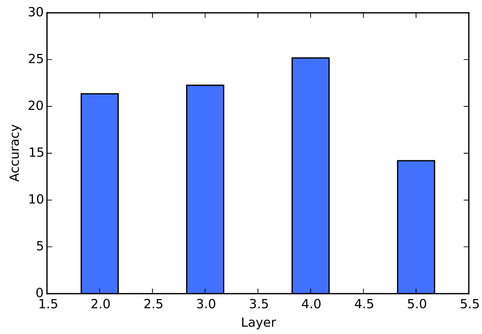

Figure 7: different layers vs accuracy with regression coefficient $\lambda = 0.001$. Size for one hidden layer is $[10]$, two hidden layers are $[20, 10]$, three layers are $[30, 20, 10]$ and four layers are [40,30,20,10].

Table 1: Layer Name and Meaning

Layer Name	Description
con1	Convolution and rectified linear activation
pool1	Max pooling
norm1	Local response normalization
conv2	Convolution and rectified linear activation
norm2	Local response normalization
pool2	Max pooling
local3	Full connected with rectified linear activation
local4	Full connected with rectified linear activation
soft_max_linear	Linear transformation to product logits

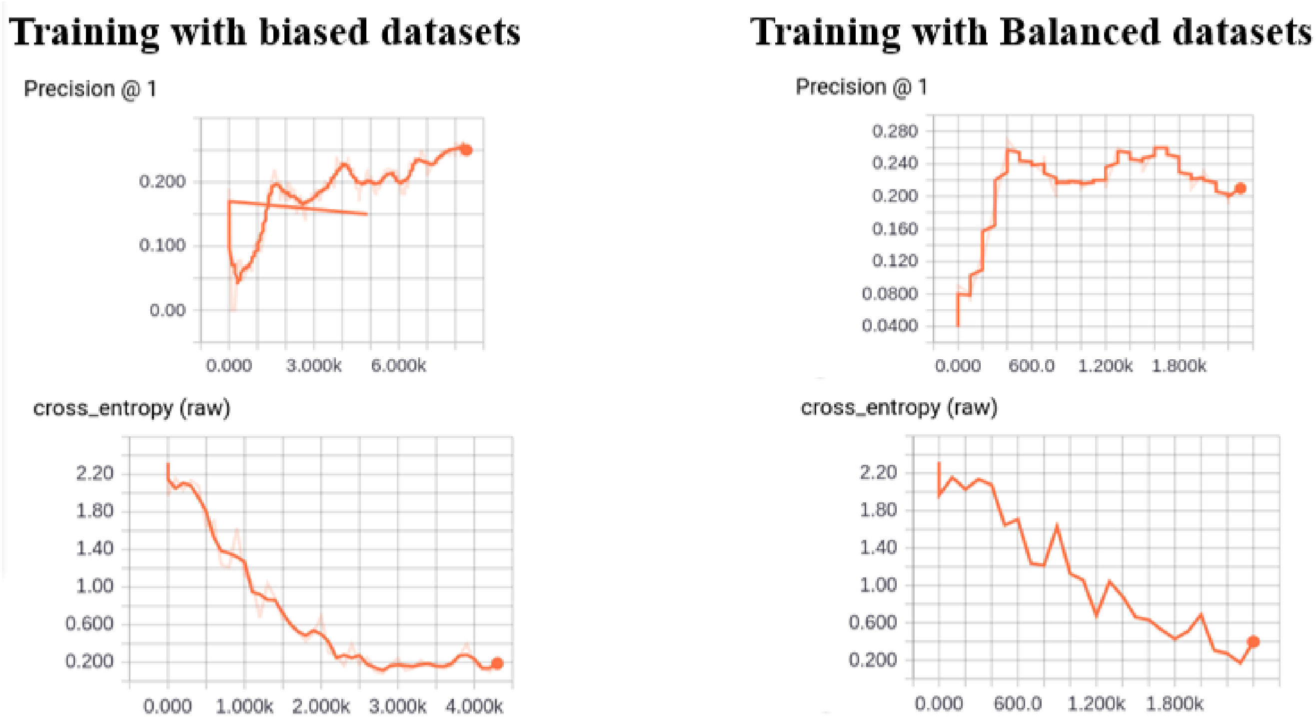

Figure 8: CNN Performance

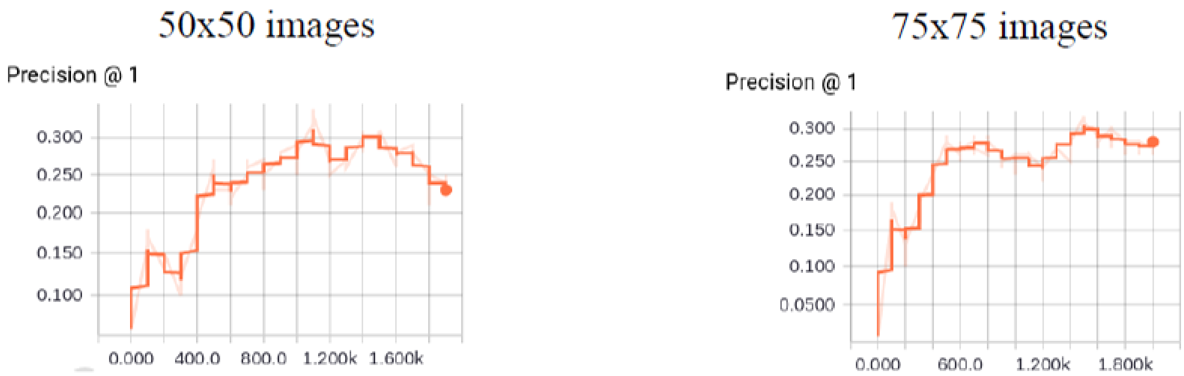

Figure 9: Performance with Larger Images Sizes

to $[30, 20, 10]$ we have the best accuracy around 25.34%.

4.2 CNN

We implement the experiments by using Tensorflow, which is a public library released by Google Inc. in 2015. All the experiments are coded by Python scripts.

It is shown in Fig. 8 that we train the model using biased and balanced datasets with image size 32×32, the predict accuracy is nearly the same. The he maximum precision is about 26%. If we train the model using the balanced datasets the learning rate is more faster than using the biased datasets.

We also conduct experiments with larger image sizes. Fig. 4.1 exhibits the performance of CNN-based method. When we increase the image sizes, the precision can be improved to be about 30%. The precision may be further improved by continuing to enlarge the image sizes, but it will take a large amount of time for training.

4.3 Discussion

We have 9 different classes in our project, so the one baseline for our predict precision is 0.11. Our model can achieve precision less than 30%, there are probably two reasons.

- The original dataset is very unbalanced so we need to duplicate part of the dataset. This option would induce overfitting to certain figures which is not the best choice. If we can have a balance dataset with no duplicate figures, the results would be more convenience.
- The advantage of DNN lies in using a huge amount of datasets. However, there are only about 1200 training data items in our project. So the performance of DNN-based classification is not very high.
- Compare with the Cifar10 project, the classification of Pinterest is not very clear. The obscurity in classification (e.g., the class of Art) makes the predict accuracy hard to improve. For example, when we check confusion matrix for most of the cases, we find 0 figures would be classified into **contexture**.
- Normally human understand concepts from concrete materials in figures (e.g. dogs, cats, chicken...etc.) and the labels may indicate a very complex concept that would not be able to extract from hidden features.

To solve the problem and to see our assumption for understanding the concrete materials is right, we re-labeled 1200 dataset into two classes as 600 with symbols (numbers and words) and 600 without symbols. Using DNN we can get 78% and CNN we can get 80% accuracy. These results indicate including new features or multi-labels may give a better prediction for the problem. Another way to approach we would suggest is to use transfer learning in the future work.

5 Conclusion

In this project, we have investigated the Pinterest image classification problem by using DNN and CNN methods. We hvae proposed engineering techniques to deal with biased training data and develop two corresponding NN models using Python on Tensorflow. Our models can achieve 30% precision for the 9-class classification problem. There are several avenues to further improve the predict precision: use larger and balanced training datasets, take more features into consideration, consider different lost function for DNN/CNN and pull layer for CNN, use new features/labels and apply transfer learning.

References

Safavian, S. Rasoul, and David Landgrebe. ”A survey of decision tree classifier methodology.” (1990).

Cover, Thomas, and Peter Hart. ”Nearest neighbor pattern classification.” IEEE transactions on information theory 13.1 (1967): 21-27.

Goller, Christoph, and Andreas Kuchler. ”Learning task-dependent distributed representations by backpropagation through structure.” Neural Networks, 1996., IEEE International Conference on. Vol. 1, 1996.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. ”Imagenet classification with deep convolutional neural networks.” Advances in neural information processing systems. 2012.

Hearst, Marti A., et al. ”Support vector machines.” IEEE Intelligent Systems and their Applications 13.4, 1998.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. ”Deep learning.” Nature 521.7553, 2015.

Available at: https://www.google.com/#q=cifar10.