Report

## Abstract

Image classification always has been a popular and basic study in artificial intelligence area. This experiment we using tow shallow learning which is support vector machine (SVM), random forests and two deep neural network one is convolutional neural network (CNN) and other is recurrent neural network (RNN) in LSTM to classify tow image datasets type. Those two dataset come derived from Microsoft dog and cat classification and CIFAR-10 that consists of 60000 32x32 color images in 10 classes, with 6000 images per class. Those two datasets are all supervised learning the dogs and cats datasets was two labels another dataset the CIFAR-10 have 10 labels, in experiment we used these four algorithms in each datasets for comparative experiments and get each algorithms performance. Before calculating we will use openCV to pretreatment each image data to for shallow learning feature extraction. During the experiment the convolutional neural networks in image recognition and processing is better than other three algorithms. Moreover, the SVM performs worse than random forest in shallow algorithms, and the specific results and reasons will be analyzed in the article.

## Introduction

需要描述一下什么是神经网络，什么是浅层学习，他们的区别是什么

Most of images recognition prefers to use convolution neural network base on it highly performance on convolution kernel which is window with space and sequence. In our research and analysis we will be take four different algorithm result in image detection area, as well as with each algorithm comparative analysis.

### RNN

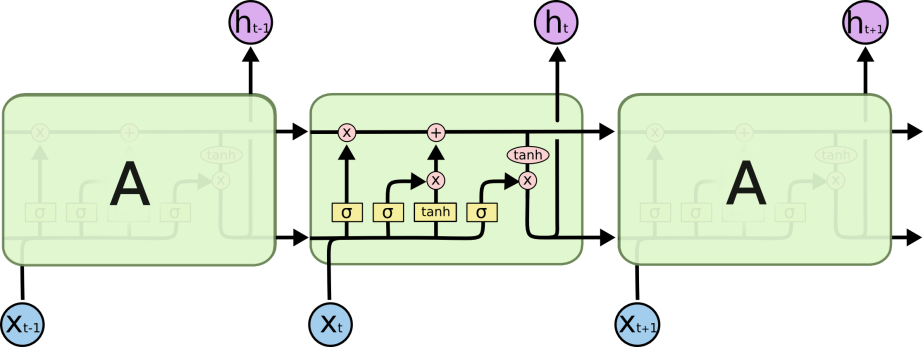
Based on our knowledge in marching learning area, recurrent neural network use a sequence to sequence way to analysis data. It is mainly used in word segmentation, part of speech tagging, named entity recognition, machine translation and speech recognition, which all belong to the scope of sequential information mining. In those areas recurrent neural network (RNN) has an unsurprisingly outstanding performance. The reason is RNN have a memory function to record information in sentence, if we detected five words sentence which means we need five layers each layer will take one words. Figure 1 show basic version of RNN network in the diagram is a layer in RNN, and t is a time step. The is the hidden state at time step t which is the most important elements in RNN call the “memory”, after formula we can see the result is determined by the hidden state of the previous layer and the input of this layer. The is a nonlinear activation function such as tanh or ReLU, because require the before layers information so generally we think it can capture information in the previous. is the output at step which means in the prediction we want to guess the next word in the sentence it would be calculate the first vocabulary, and using sorftmax function to convergence result .



Figure 1

（<http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/>）

At same time, in our experiment we try to using one of types RNN which call the LSTM to predict images. LSTM proposed by Hochreiter & Schmidhuber in 1997, it has been specially designed to avoid the long-term dependency problems that occur in RNN. Furthermore, LSTM was different between traditional RNN a simple neural network, in the LSTM each neural have four interaction layers. In the top of single neural there is a main procession to ensure the integrity of the information, and LSTM using a special way call the “gate” to control and select data through which include sigmoid layer and pointwise multiply operation.



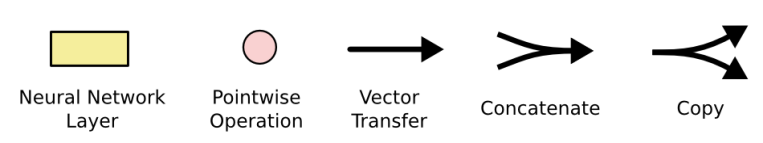
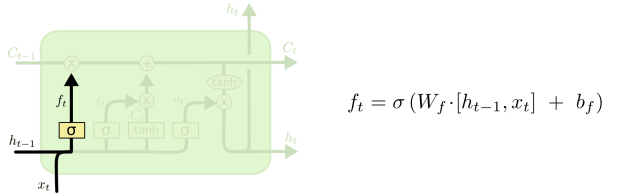
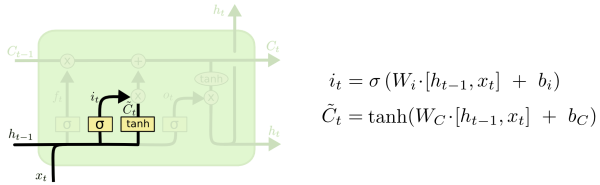


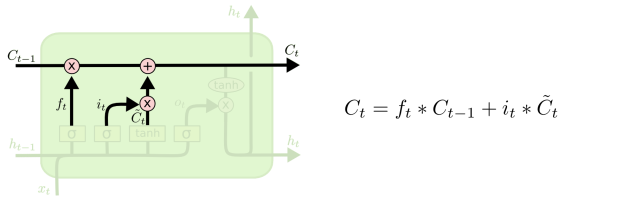
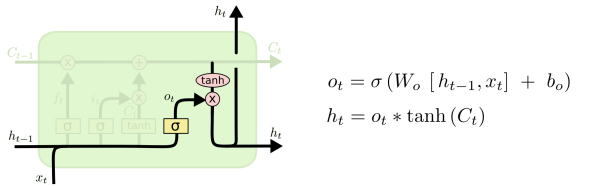
Figure 2

(<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>)

First layer is a “forget gate layer” that is a sigmoid function to compress the data from 0 to 1 each number in will be filter, 0 will be discard and 1 is completely keep. Second layer is “input gate layer”, this layer also using input parameters and to create a vector will be using on next step. ’s result will be multiply older forget value and add new data. Finally, we get next layer which is the information we wanted to output by using, to get sigmoid function and multiply tanh.

Step 1 Step 2

Step 3 Step 4

Figure 3

In dogs and cats dataset is a supervised learning, we provide two labels image and split it to two each training and testing dataset is 12400 and 12546. We designed a 200 units LSTM network, and considering the convergence of gradient descent, we used 64 as the batch size, however the finally the model accuracy is 60%. Compare with 10 labels CIFAR-10 datasets the accuracy of the model was increased by ten percent. The performance of the same model on two different data sets has a huge distinction that we think it is not only because of the different number of images in the two data sets, but also because of the different number of tags in the images.

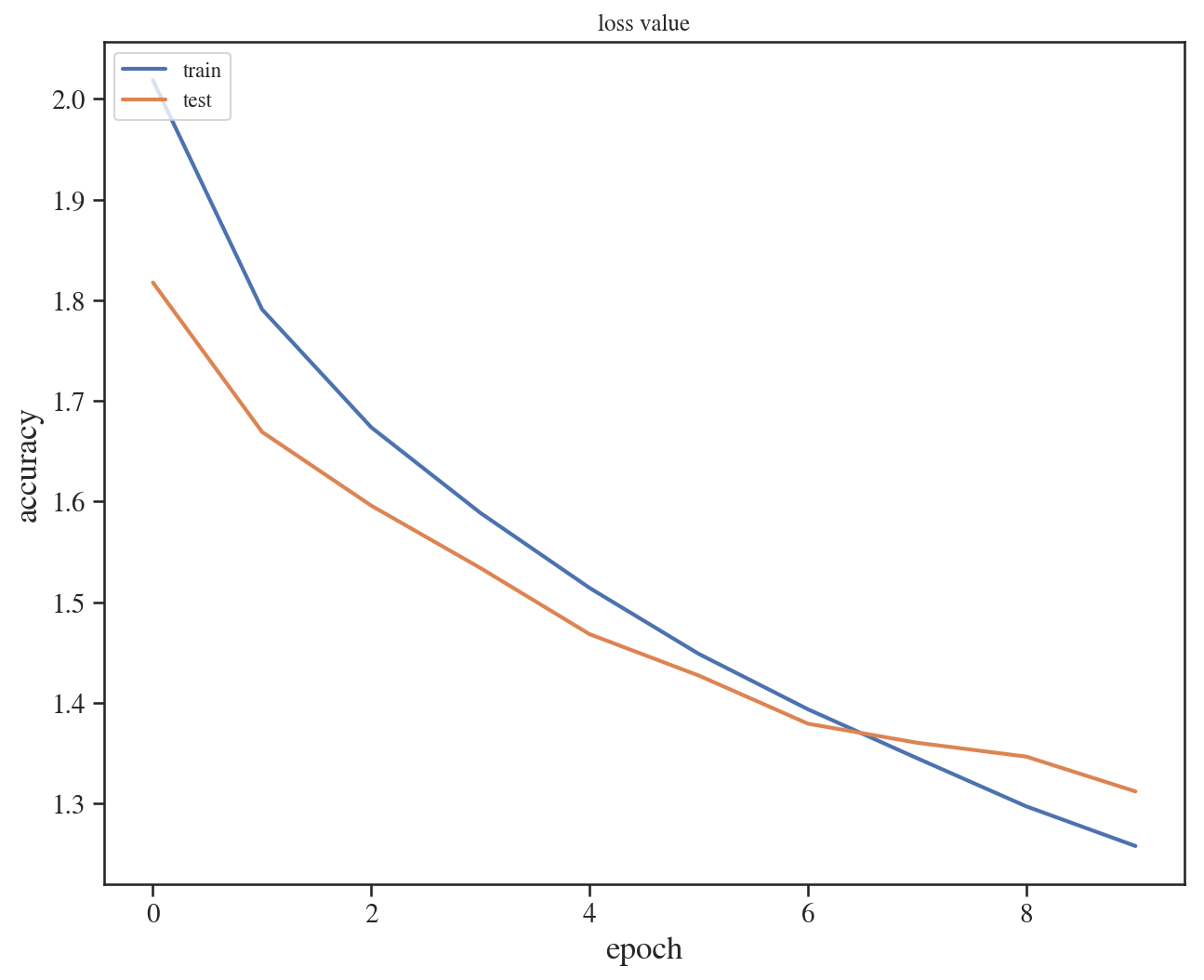
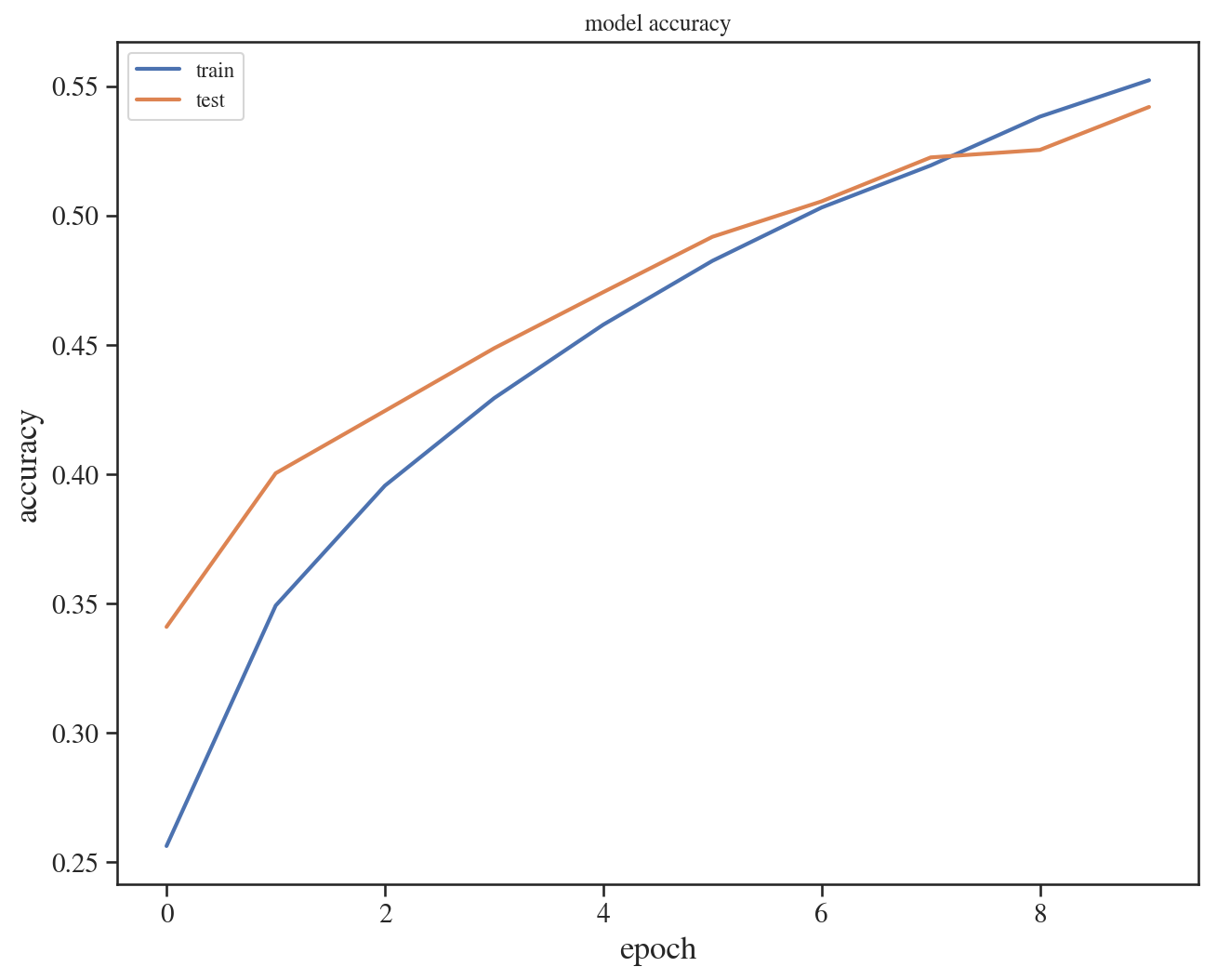


Figure 4

### Random forest

Random forest is an ensemble learning method and it is usually used to deal with classification problems. Ensemble learning occurs when other machine learning methods do not perform well. It combines multiple machine learning methods together to solve a single prediction problem. Random forest is one of them, it composed of a large number of decision trees method together and each decision tree study by oneself make a single prediction, combine all the decision tree prediction result together become to random forest. There is lots of advantage in random forest method such as high accuracy; the randomness makes it difficult to over fit the random forest; the randomness makes the random forest have good anti-noise ability and so on.

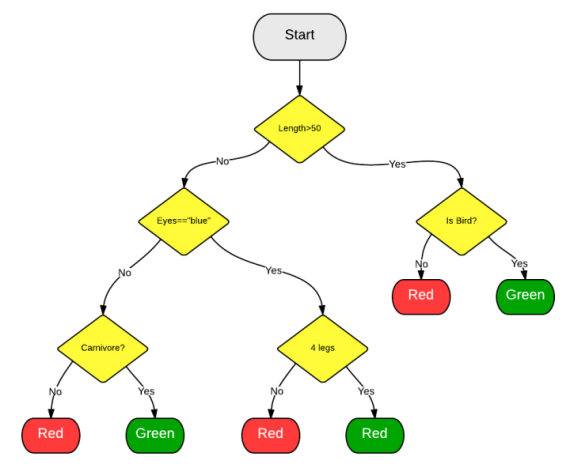


Figure 5 Figure 6

(<https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d>)

(<https://zhuanlan.zhihu.com/p/23614800>)

The figure 6 a single decision tree model prediction process and the random forest algorithm will automatically create a random group of decision tree. Since these trees are randomly generated, most of them (even 99.9%) don't make sense for solving the classification or regression problem. However, when each decision tree has a prediction label from top to bottom, all the predicted results will be sorted out by the random forest model to get the final predicted results. The error rate will be offset from the quantitative random forest model, and the optimal decision tree will be extracted.

In our experiment we start to use OpenCV as the sample feature extraction method for our training, and for the image we extract a 3D color histogram from the HSV color space using. Then we reduce the dimension of the image matrix by PCV, the image matrix are placed into random forest model, we got the finally result about dogs and cats datasets accuracy was 66% and for CIFAR-10 datasets the finally accuracy only 27%. The random forest method does not perform well in the classification of a large number of images in our experiemce.