University of Aizu, Graduation Thesis. March, 2021

Classifying Television Commercials applying Neural Networks with Evaluation of Error Functions

Ryota Moriya s1250103 Supervised by Prof. Kazuyoshi Mori

Abstract

In this paper, we present a development of a system to classify television commercials following categories. Recently, television commercials are familiar to us living in society. We consider that investigating television commercials can be useful for social analysis because these reflect people's trends, culture, and society.

Thus, we decided to develop a system to classify television commercials following some categories. We attempt to apply Neural Network for the development.

On the other hand, in many applications of classifier learning, training data suffers from label noise. Label noise makes the generalization ability of applications lower. However, label noise is not clear and easy to observe. To observe label noise, we focus on error function through learning.

1 Introduction

Recently, neural networks have exhibited very impressive performance in many classification problems. On other hand, investigation of television commercials (CMs) is a powerful method for social analysis. Such a system that classifies CMs automatically is helpful for investigators.

Every CM is a span of television programming produced and paid for by an organization and sophisticated to become the maximum effect to its cost and efficiently attract to people motivation. After watching each CM, people take economic actions. Thus, we consider that CMs reflect trends, culture, and society.

Actual investigating, it is needed for investigators to classify CMs following some categories. However, it is hard for a human to process such labeling work.

In this paper, we present a classifier that classifies CMs automatically following some categories applying neural networks. We refer to the system as a CM classifier. Besides, we examine the influences of label noise on the CM classifier to develop a better classifier focusing on error function during its learning. Because we consider that the actual difficulty of labeling works seems to be a large reason against label noise in CM classifier learning.

Related Work

The developments of CM classifier applying neural networks have been examined in some approaches. The relation of label noise in classifier learning and loss function is shown in some researches.

Convolutional neural networks were utilized for the development of the CM classifier. ReLU applies as an activation function in hidden layers. There are some empirical results of low accuracy of CM classifiers. Detailed surveys are available in [5][6].

Label noise is one significant reason for the lower generalization ability of many applications of classifier learning. When the class labels in the training data are noisy (i.e., may be incorrect) then it is referred to as label noise. Human labeling errors, measurement errors, and subjective biases of labelers are among the reasons for label noise although some are not arbitrary.

In the case of classifying CM, designing the category for classification equals the summarization of CMs all over the world, which number is very large. Besides, labelers classify CMs for learning following the design category. Thus, the work is a complicated process. Through the process, class labels become noisy.

The use of some loss functions in neural networks derives each different result. Detailed surveys from theoretical and empirical results are written in [1].

The label noise is not clear and easy to observe. In this paper, we discuss label noise with results through some CM classifier learnings implemented each different error function. Preliminaries

In this section we introduce some technical words and software related to actual our development.

1.1 Neural Network

Neural networks are computing systems modeled at the human brain and nervous system. It is often utilized in the field of pattern recognition such as character recognition and speech recognition. Neural networks consist of the input layer, hidden layer, and output layer mainly.

Convolutional neural networks (CNNs) are one type of neural networks with convolutional layers and pooling layers in the hidden layer. Recently, CNNs are often utilized in the field of image recognition. 2D CNNs extract features from spatial dimensions by performing convolutions, thereby the motion information encoded in multiple adjacent pixels. CNNs are also applied to video recognitions and classifications. 3D CNNs extract features from not only spatial dimensions such as 2D CNNs, but also the temporal dimension.

1.2 TensorFlow and Keras

Every TensorFlow and Keras is an open-source library for machine learning. TensorFlow is a machine learning platform that lets us to use basic machine learning. Keras is running on top of TensorFlow. In the development in this paper, we use these.

1.3 OpenCV

Open Source Computer Vision Library (OpenCV) is also an open-source library [2]. This is implemented in many functions for dealing with images and videos on a computer. In this paper, we process original CM data to the appropriate format using this library.

1.4 Colaboratory

Colaboratory, or 'Colab' for short, is a product from Google Research. Colab allows anybody to write and execute arbitrary Python code through the browser, and is especially well suited to machine learning, data analysis and education. More technically, Colab is a hosted Jupyter notebook service that requires no setup to use, while providing free access to computing resources including GPUs.

2 CM classifier development

In this section, we introduce the development of a CM classifier.

We developed a system that outputs the one suitable category for input CM. We designed “food”, “car” and “cosmetic” as the categories for classification in supervised learning in advance. Note that “other” is similarly one category for the classification. Consequently, four categories are designed.

2.1 Preparation of Learning Data

We had collected different 254 CMs, which were 15 seconds although there are various time CMs in general. We classified these CMs with the designed categories and attached the corresponding class label to these CMs. The class labels were identified as the correct answers in supervised learning.

We had divided these CMs in 4:1 as each training data and test data. Finally, we had resized these CMs due to computer resources however we confirmed that these CMs were classifiable also. Consequently, the original CMs, which have spatial dimensions expressed with 1920 width pixels and 1080 pixels were processed to 80 pixels and 45 pixels respectively.
Any original frame size was reshaped to 30 by sampling at appropriately equal intervals. RGB was utilized for color information. The image of visualized one video as the learning data is shown in Figure 1.

Figure. 1: One example of CM for the classifier learning. Every CM is expressed as one sequence of 30 images.

2.2 Networks and Learning

All networks utilized Rectified Linear Unit (ReLU) as an activation function in the hidden layers and have SoftMax layer at the output with the size of the layer being the number of category classes. The hidden layer consists two of 3D convolutional with pooling layers for extracting features from the spatial and temporal dimensions. The visualization of the Neural Networks is shown in Figure 2.

The learnings are independently for every error function. CCE, MSE, and MAE are utilized for learning. All networks are trained through backpropagation with momentum term and weight decay. We applied stochastic gradient descent with 32 batch sizes and 128 epochs.

2.3 Results

The results of the learnings are shown in table 1. In Table 1, we see the values of accuracy for test data using each error function.

Although accuracies of CCE and MSE achieved over 0.5, we concluded that our CM classifier achieved low generalization ability. However, frankly speaking, the accuracy of CCE achieved 0.51 and the accuracy of MSE achieved 0.53. The value of MSE is a little higher than that of CCE. A detail of the leanings is shown in figure 3.

Figure. 2: Visualization of the Network Networks for CM classifier

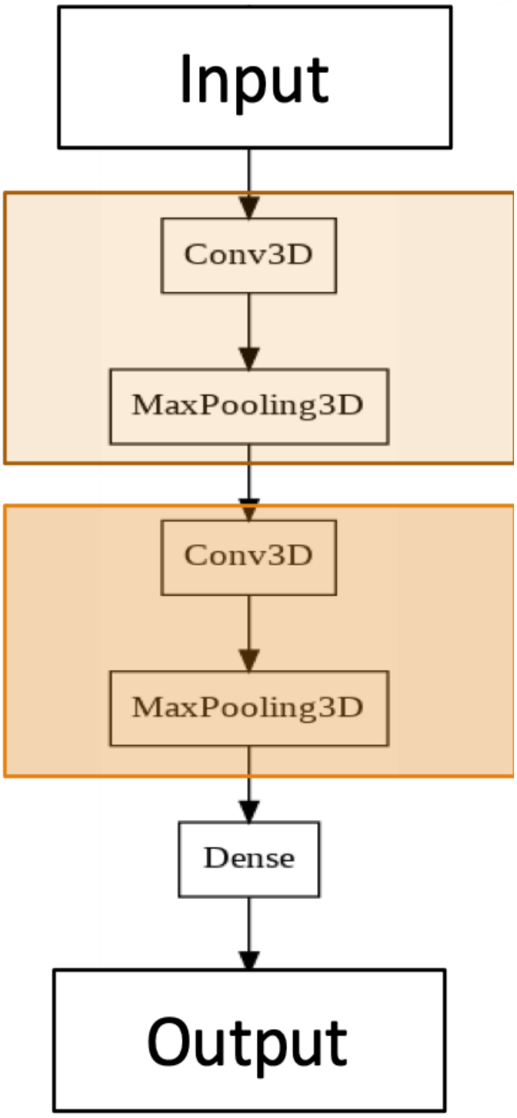

Table 1: Test Accuracy for error functions

Loss	Accuracy
CCE	0.515
MAE	0.406
MSE	0.531
Average	0.484

3 KTH classifier development and Label noise experiment

In this section, we introduce the development of a novel classifier and an experiment about label noise with it.

Our CM classifier showed low accuracy. Therefore, we researched the reason for the view of label noise. We consider that the experiment with another video classifier relates to a CM classifier as one video classifier.

We selected one dataset which seemed to be lower influences of label noise and developed a novel classifier.

3.1 KTH classifier development

KTH dataset is one dataset that is the most widely utilized in motion recognition researches [3][4]. These data set categories are simpler and clearer compared to the ones of CM. We expected that the KTH classifier is less influenced than the CM classifier.

The videos of the KTH dataset had been already classified for various motions. We utilized the four categories such as “walking”, “running”, “boxing” and “clapping” for the learning and the total is 399 finally. We had divided these videos in 4:1 as each training data and test data and resized 80x45x30 with RGB color information for 3D convolutions.

The networks of the KTH classifier are the same as the ones of the CM classifier. The learnings are also independent for CCE, MSE, and MAE. The networks are trained through backpropagation.

3.2 Label noise experiment using error functions

We introduce the actual procedure for the experiment to observe the influences of label noise against the classifier learning.

First of all, we developed a good classifier for good classification. Next, we developed also other systems, which learned under intentional label noise. Note that we developed a good classifier to observe the influences of intentional label noise initially. We refer to this good classifier as the classifier under no intentional label noise.

We intentionally changed some class labels attached training data to the wrong value for learnings under intentional label noise. The wrong value was assigned randomly. The classifies learned under the conditions changed its class labels to each 0%, 20%, and 30%. As a result, the three classifiers suffered from intentional label noise. When the class labels were changed 10% for all training data, it was referred to as 10% label noise. The same applied to 20% and 30% label noise.

Note that when there was no intentional label noise, it was referred to as 0% label noise.

3.3 Results

The result of the experiment is shown in table 2. We see accuracies for each error function under each intentional label noise.

Under 0% noise, the average accuracies in three error functions achieved 0.78. Thus, we conclude that this system achieved enough accuracy as a video classifier.

Comparing accuracies of each error function, the value of CCE achieved 0.8 and the accuracies of MSE and MAE achieved 0.78. The value of CCE is higher than other values.

Under 10% noise, the average accuracies achieved 0.71. Moreover, the average under 20% noise achieved 0.64, and under 30% noise achieved 0.63. We see that more label noise is less the accuracy. Looking at the values of CCE, the accuracy has already achieved a little less than 0.74 of MSE. Under the noisier condition, CCE was not the highest in the three error functions. A detail of the leanings is shown in figure 4.

Table 2: Test Accuracies for each error function under each intentional label noise

Loss	Accuracy			
-----------	***0%***	***10%***	***20%***	***30%***
CCE	0.80	0.73	0.60	0.59
MAE	0.78	0.66	0.68	0.71
MSE	0.78	0.74	0.66	0.59
Average	0.786	0.710	0.646	0.63

4 Conclusion and Future Works

In this paper, we introduced the development of the CM classifier. However, our CM classifier did not show good results. We also developed a KTH classifier with little original label noise and processed the experiment to observe influences of label noise to another video classifier.

We obtained that higher label noise was the lower generalization of the classifier. Moreover, we also obtain that learnings of CCE under the noise label were not superior to MSE and MAE. In Table 1, we see the accuracies of CM classifiers achieved low and the accuracy of CCE achieved not higher of the three error functions. Hence, we conclude that our CM classifier suffered from label noise.

In future work, we attempt to the extension of category. At this time, we designed the category such as “food”, “car” and “cosmetic” simply. We can design a more complicated category also. For example, “Frequent appearance of actress”, “Advertisement for medical product” and so on. More detailed and much information lead to more advanced classification.

References

[1] Aritra Ghosh, Himanshu Kumar and P. S. Sastry, “Robust Loss Functions Under Label Noise for Deep Neural Neural Networks”, 2017, https://ojs.aaai.org/index.php/AAAI/article/view/10894

[2] “OpenCV documentation Index”, http://docs.pencv.org/

[3] Dong Han, Liefeng Bo and Cristian Sminchisescu, “Selection and context for action recognition”,2010, https://ieeexplore.ieee.org/abstract/document/5459427/authors#authors

[4] Amir Nadeem, Ahmad Jalal and Kibum Kim , “Human Actions Tracking and Recognition Based on Body Parts Detection via Artificial Neural Network”, https://ieeexplore.ieee.org/abstract/document/9055951/authors#authors

[5] Ryu Nagabayashi, “Classifying Television Commercials by Convolutional Neural Network with Evaluation of Activation Functions”, 2019

[6] Takahiro Murakami, “Classifying Television Commercials by Convolutional Recurrent Neural Networks”, 2020

Figure. 3: Plots of training and test accuracies for each error function during the learning of CM classifier

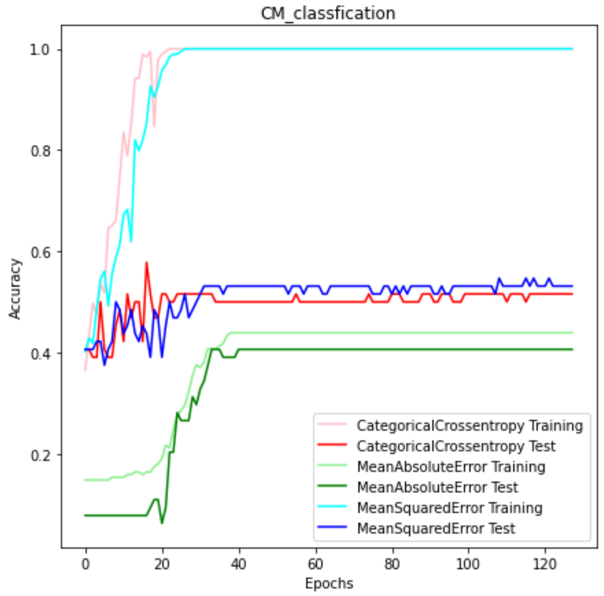

Figure. 4: Plots of training and test accuracies for each error function during the learning of CM classifier under each intentional label noise