Classifying Television Commercials by Convolutional Neural Network with Evaluation of Activation Functions

畳み込みニューラルネットワークによる **CM** 分類と活性化関数の評価

s1230179 Ryu Nagabayashi, Supervisor: Prof. Kazuyoshi Mori

Ⅰ. Motivation and Goal

A television commercial (CM) provides various information on products and services. People watch television on a daily basis, and CMs have influences to people and the social culture. We consider that investigating many CMs is useful for social analysis, especially business purposes. Therefore, we have decided to develop a semiautomatic system by convolutional neural network for classifying CMs.

In the neural network, a function called activation function[1] is used. There are many kinds of activation functions and they often influence to significant results of the neural network. We consider that evaluating activation functions exercise a positive effect for making the system more effective.

Figure 1: CM image labeled "food"

Figure 2: CM image labeled "car"

Figure 3: CM image labeled "cosmetic"

Ⅱ. Approach

We have used convolutional neural network (CNN)[1] in order to classify a television commercials. CNN is a kind of deep learning. It has several layers including convolution layers and pooling layers. It is often used for image recognition.

In many cases, images are inputs one by one to CNN. In this study, we implement CNN with video inputs by using several consecutive images for input. This will enable us to obtain a temporal feature of CMs.

In order to identifying CMs, networks[2] we use are Open Source Computer Vision (OpenCV)[3] and TensorFlow[4].

Ⅲ. Current Results

In order to training CNN, we have so far collected 68 of various CMs, in which 53 videos are used to train CNN, and 15 videos are used to test it. These CMs are labeled a category, such as “food” (Figure 1), “car” (Figure 2), and “cosmetic” (Figure 3).

We have implemented a preliminary CNN (Figure 4) that can classify CM videos into three categories "food", "car", or "cosmetic." It consists of 6 convolutional layers and 2 fully connected layers. Currently, ReLU[1] and softmax[1] are used as activation functions. The second, fourth, and sixth convolution layers are followed by the pooling layer. We have here used max pooling. In the input layer, one CM should play a role of input, where a CM is a video. Even so, in the current system, we input 30 successive images from one CM, rather than a CM video itself. All of CMs are originally recorded as 1920 x 1080 resolution, but for the computer source limitation, we have resized them to 80 x 45.

The preliminary results of system evaluation are shown in Figure 5. The current accuracy rate is not so high, in which we do not satisfied yet. We have considered that further consideration of the activation functions would be necessary. Moreover, increasing the number of CMs used for training will result in better accuracy rate.

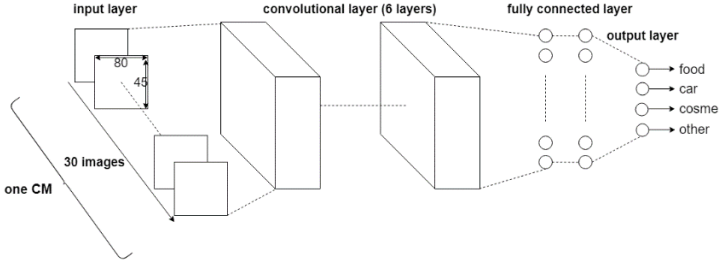

Figure 4: Convolutional Neural Network

```
food CMs
-------------------------------------
Input: food1.mp4   result: food
Input: food2.mp4   result: food
Input: food3.mp4   result: food
Input: food4.mp4   result: food
Input: food5.mp4   result: cosme
-------------------------------------
accuracy rate: 0.8
-------------------------------------
car CMs
-------------------------------------
Input: car1.mp4    result: food
Input: car2.mp4    result: food
Input: car3.mp4    result: car
Input: car4.mp4    result: food
Input: car5.mp4    result: food
-------------------------------------
accuracy rate: 0.2
-------------------------------------
cosme CMs
-------------------------------------
Input: cosme1.mp4  result: cosme
Input: cosme2.mp4  result: car
Input: cosme3.mp4  result: cosme
Input: cosme4.mp4  result: cosme
Input: cosme5.mp4  result: food
-------------------------------------
accuracy rate: 0.6
-------------------------------------
overall accuracy rate: 0.533
```

Figure 5: Test result

Ⅳ. Schedule

As we have implemented a preliminary system. We will evaluate a number of activation functions such as ReLU, step function[1], and sigmoid function[1], in order to have more precise classification. We also need to evaluate number of layers, sizes of layers, and so on. As a future work, we intend to improve the accuracy by using additional data and discuss CNN by changing activation functions.

Ⅴ. References

[1]原田達也, “画像認識”, 講談社, 2017.
[2]Hiromi Hirano, “C でつくるニューラルネットワーク,” パーソナルメディア株式会社, 1991.
[3]“Neural Networks — OpenCV 2.4.13.1 documentation,” 2014, http://docs.opencv.org/2.4/modules/ml/doc/neural_networks.html
[4] “MNIST For ML Beginners,” Dec. 2016,https://www.tensorflow.org/versions/r0.11/tutorials/mnist/beginners/index.html#the-mnist-data