

Classifying Television Commercials by Convolutional Recurrent Neural Networks

畳み込みリカレントニューラルネットワークによるCM分類

s1240099 Takahiro Murakami, Supervisor: Kazuyoshi Mori



1. Motivation and Goal

Television commercials (TVCMs) provide various information about products and services to sell them. Thus, TVCMs reflect the social situation and trend at that time. These often affect people and social culture. We consider that investigating many TVCMs is effective for social analysis, especially business purposes. Therefore, we have decided to develop a semiautomatic system by convolutional neural networks (CNNs) and recurrent neural networks (RNNs)[1] for classifying TVCMs. Thus, we decided to develop a semi-automatic system for classifying TVCMs. A system using a convolutional neural network requires a large amount of data to be input at one time. By incorporating a regression type neural network into this system, a system with less data to be input at a time will be developed. By using CNN and Long Short Term Memory(LSTM), it is considered that TVCMs can be recognized by moving images.

2. Approach

To classify TVCMs, CNNs and RNNs are used. CNN and RNN are a kind of deep learning. RNN(Figure1) has the feature that it can be referred to by subsequent data while retaining data at a certain point. For this reason, RNNs are often used for speech recognition and language processing[2]. CNN(Figure2) has several layers, including several convolutional layers and a pooling layer. CNN is often used for image recognition. First of all, this research uses 15 images as input. CNN is used to predict the classification of an image. RNN is used for the output. Do this for all 15 images. The networks used to identify CMs [3] are based on Open Source Computer Vision (openCV) [4] and Tensor Flow [5].

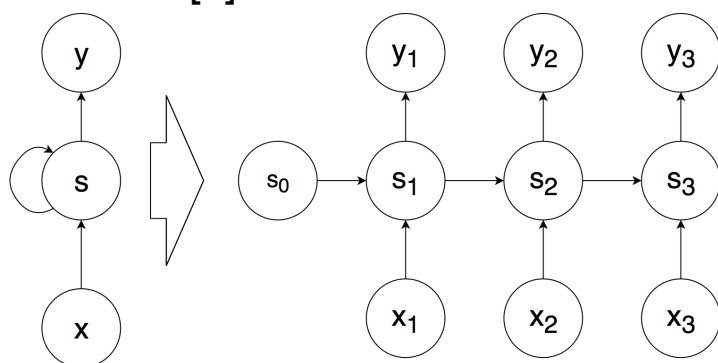


Figure1: Recurrent Neural Network

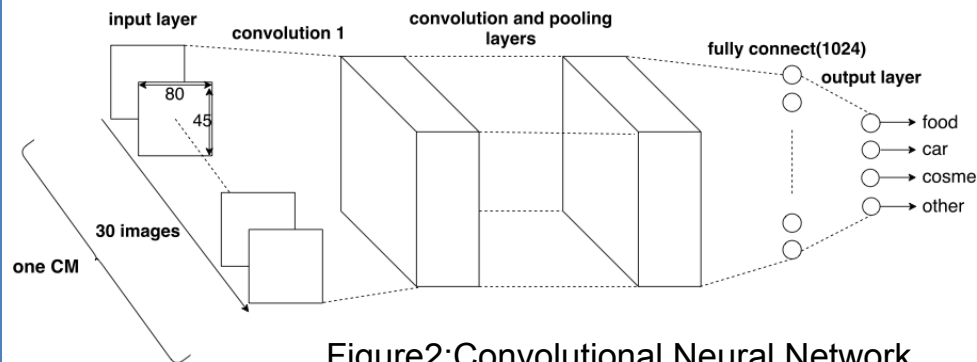


Figure2:Convolutional Neural Network

3. Current Results and Status

In order to training CNN, we have so far collected 68 of various TVCMs, each of which is labeled a category, such as “food”, “car”, and “cosmetic”. We have implemented a preliminary CNN that can classify TVCMs into three categories "food", "car", or "cosmetic." It consists of 6 convolutional layers and 2 fully connected layers. One CM should play a role of input, where a TVCM is a video. In the current system, we input 30 successive images from one TVCM. In addition to this, RNN should be used to reduce the amount of data entered at one time.

4. Remaining Tasks and Tentative Schedule

Since the TVCM classification system by CNN is implemented, our next step is to implement RNN and LSTM and then we make evaluation for more accurate classification.

5. References

- [1]Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016.
- [2]原田達也, “画像認識”, 講談社, 2017.
- [3]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.
- [4]“Neural Networks — OpenCV 2.4.13.1 documentation,” 2014, http://docs.opencv.org/2.4/modules/ml/doc/neural_networks.html
- [5] “MNIST For ML Beginners,” Dec. 2016, <https://www.tensorflow.org/versions/r0.11/tutorials/mnist/beginners/index.html#the-mnist-data>
- [6]坂和正敏,田中雅博“ニューロコンピューティング入門”,森北出版株式会社,1997.