Based on the Video Multimedia Technology, A Learning Analysis Application Research in the Wisdom Classroom

Shao-hui CHEN¹, Lei XIANG¹ and Rui JIN²

¹National Engineering Research Center for E-Learning, Central China Normal University, Wuhan, China 430079
²Hangzhou Normal University, Hangzhou, China 310018

Keywords: Multimedia technology, Convolutional neural network (CNN), Wisdom classroom, Learning analysis.

Abstract. In the traditional teaching activities, teachers according to students' actual level, combined with their own experience and take the appropriate method, undertake to the student learning situation analysis, can avoid empirical teaching design and teaching material as the center of teaching. However, a gap exists between the theory and practice of learning analysis. How to get more and more accurate information about the students learning from the big data occurring in the process of teaching, increase teacher’s effective attention to specific student, is an effective way to improve the teaching effect. How to effectively and accurately focus on students’ classroom learning state, such as attention, listen, speak, think, classroom practice, emotional, physical, and so on and so forth, to improve teaching efficiency. This has important significance.

In this paper, we use the model of deep learning convolution neural network technology to classify many objects (students) appeared in the smart classroom video. And analysis and identify the each target’s (a student) behavior (focus on listening to lectures, hands up, stand up, writing, sleeping on table, etc.), judge and record each goal’s (students) statement of learning, to provide accurate, fine and explicit learning analysis basis for teachers.

Introduction

Deep learning is very popular in the field of machine learning in recent years. The recognition effect of convolutional neural network (CNN), as a representative of deep learning network than the traditional neural network has a huge increase. The method is a kind of end-to-end recognition method, and don't need the manual design features. It has attracted a lot of people to study, and have already succeeded in some areas of computer vision. It has a certain translation invariance and scale invariance, and its calculation method and the mammalian visual system has a lot of similarities.

The Development History of CNN

The section headings are in boldface capital and lowercase letters. Second level headings are typed as part of the succeeding paragraph (like the subsection heading of this paragraph). In 1959, Two scientists, Hubel and Wiese, found cat's vision system is hierarchical in the visual experiments of cat. This kind of classification can be as the process of iteration and abstract step by step: pupils receive pixels, directional selectivity cell abstract the edge, the edge further abstract shape, the shape further abstract what objects are obtained. The advanced features are combination and abstraction of lower features, the more advanced features can more reflect the semantic information of human beings. Then the researchers put forward a new word to name this model - "deep learning model". In 1980, Kunihiko Fukushima proposed ‘Neocognitron’ first introduced the concept of CNN,This is the first deep learning model. Later, in practice and theory analysis, many scholars made a significant contribution to the development for CNN. In 1988, LeCun introduce BP algorithm to CNN. In 2003, Behnke wrote a book on CNN, the CNN are summarized. That same year, CNN is improved further by Siresan etc, and implements the GPU version of it. later, they use the CNN frame to do experiments on multiple image database, have been making the best results.
The Network Structure of CNN

Convolutional neural network (CNN) is a multi-layer neural network. The basic structure of it consists of the input layer, convolution (conv) layer, sampling (pooling) layer, whole connection layer and the output layer (classifier). A classic CNN structure is shown in figure 1.

![Figure 1. The basic frame of CNN.](image)

The Introduction of CNN. Convolution transformation action between one layer and another layer in CNN is a process of feature extraction. Each layer of CNN is composed of multiple 2 dimension planes. Each plane is a feature map after the characteristics extract processing (fm). The original image of input layer and each feature extraction image in the middle layer (convolution layer) are followed by the second feature extraction computation layer (a subsampling layer). This typical secondary feature extraction structure makes CNN have a certain deformation tolerance to the input data. There are several convolution and sub sampling layers. The convolution process is as follows:

1. Convolution process: Using a trainable filter (fx) to do convolution in the input image (or a layer of feature map), and then add offset (bx), get convolution layer(Cx);

2. Subsampling process: sum four pixels in each neighborhood to get a pixel, weighted by scalar WX + 1, then add the bias bx + 1, and then through a sigmoid activation function, get a contracted image (Sx + 1) about a quarter of feature mapping.

The process of Convolution and subsampling layer is shown in figure 2

![Figure 2. The process of convolution and subsampling.](image)

Convolution layer is feature extraction layer. The convolution operation can strengthen the original signal characteristics, at the same time reduce the noise. Sub sampling layer is used for the secondary feature extraction, can be regarded as a fuzzy filter.

1. The full connection layer (F layer) is equal to the HiddenLayer in the multilayer perceptron (MLP) of traditional neural network, and fully connected with the before layer. The calculation process is the result of the before layer’s output multiplied by the weight vector, plus an offset, and then put it to the sigmoid function.

2. The output layer, classification layer, consisting of European Radial basis Function (Euclidean Radial Function) units. Each category corresponds to a unit. The output layer using Logistics regression, calculating the probability of types of input sample.
CNN Used for Classification of Classroom Learning Behavior

The Construction of Used CNN

![CNN Framework](image)

Figure 3. The CNN framework of identifying classroom learning behavior.

CNN on classroom learning behavior classification recognition can be divided into three layers: input layer, two convolution layer, a full connection layer, an output layer. Input image size is 268×268 pixels, convolution kernels is 5×5, feature mapping image size is 264×264 pixels, the number of the output classification is 5 classes.

The Preparation of the Test

**The Preprocessing of the Input Image.** The processing image is grayscale image. Using filtering technology to eliminate image noise and the normalized processing technology to resize to 256×256 grayscale images.

**The Definition of the Output.** Class number was expressed in digits.

Table 1. Pattern classification results of CNN.

<table>
<thead>
<tr>
<th>Classification</th>
<th>The Expression of the CNN Output</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>class 1</td>
<td>00001</td>
<td>the first category</td>
</tr>
<tr>
<td>Class 2</td>
<td>00010</td>
<td>the second category</td>
</tr>
<tr>
<td>Class 3</td>
<td>00100</td>
<td>the third category</td>
</tr>
<tr>
<td>Class 4</td>
<td>01000</td>
<td>the fourth category</td>
</tr>
<tr>
<td>Class 5</td>
<td>10000</td>
<td>the fifth category</td>
</tr>
</tbody>
</table>

**The Definitions of CNN Training and Testing Image Data Sets.** Image library contains 5 categories, 150 images. Each category contains 30 images. Each category contains 20 images as training data, the remaining 10 images as test data.

**The Experimental Steps**

1. Specify the path and configuration file of the training image set.
2. Specify the path and configuration file of the test image set.
3. Conduct the image grey value and the size normalization: 268×268.
4. The experimental image set were randomly divided into two: training image set (5×20 images) and test set (5×10 images), the training image set (5×20 images) is directly input into neural network training.
5. Test image set (5×10 images) is directly input into the neural network testing.

**The Experimental Results**

CNN after training, the prediction accuracy of images classification is showed in the following table:

Table 2. CNN corresponding forecast results.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Identify the correct number images /total number of images</th>
<th>The prediction accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN</td>
<td>31 / 50</td>
<td>61%</td>
</tr>
</tbody>
</table>
The Experimental Conclusion

The recognition accuracy of three layers CNN for Learning behavior in the classroom scene video reaches 61%. If we continue to improve the structure of the CNN and increase the number of training data set can improve the accuracy of recognition.

CNN Used for Multi-objective Classification in the Classroom Video

The Construction of the Used CNN

![Diagram of CNN structure](image)

Figure 2. Multi-target classification framework (CNN) classroom situations.

CNN multi-objective classification recognition in classroom learning can be divided into three layers: input layer, two convolution layer, a full connection layer, an output layer. Input image size is 268×268 pixels, convolution kernels is 5×5, feature mapping image size is 264×264 pixels, the number of the output classification is 5 classes. To classify five target tracking identification.

The Preparation of the Test

The Preprocessing of the CNN Input Image. To use 150 images for image gray processing. Using filtering technology to eliminate image noise. And the normalized treatment for 256 * 256 grayscale images.

The Definition of the Output. The output was expressed in digits.

<table>
<thead>
<tr>
<th>Classification</th>
<th>The Output of the CNN</th>
<th>The Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>00001</td>
<td>The first goal</td>
</tr>
<tr>
<td>Class 2</td>
<td>00010</td>
<td>The second goal</td>
</tr>
<tr>
<td>Class 3</td>
<td>00100</td>
<td>The third goal</td>
</tr>
<tr>
<td>Class 4</td>
<td>01000</td>
<td>The fourth goal</td>
</tr>
<tr>
<td>Class 5</td>
<td>10000</td>
<td>The fifth goal</td>
</tr>
</tbody>
</table>

The Definition of CNN Training and Testing Image Data Set. Image library contains 150 images, including five students common actions in class. Each student has 30 images. Select 140 images as training data, the remaining 10 images as test data.

The Experimental Steps

1. specify the path and configuration file of the training image set.
2. specify the path and configuration file of the test image set.
3. conduct the image grey value and the size normalization : 268×268.
4. The experimental image set were randomly divided into two : training image set (5×28 images) and test set (5×2 images), the training image set (5×28 images) is directly input into neural network training.
5. Test image set (5×2 images) is directly input into the neural network testing.
The Experimental Results

After training, the prediction accuracy of CNN are classified in the following table:

Table 4. The prediction results of target classification.

<table>
<thead>
<tr>
<th>methods</th>
<th>accuracy</th>
<th>the correct number/total number of images</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN</td>
<td>70%</td>
<td>7 / 10</td>
</tr>
</tbody>
</table>

The Experiment Conclusion

The recognition accuracy Of three layers CNN for multi-objective classification and recognition in classroom scene by processing the image is of 70%. Continue to improve the structure of the CNN can improve the accuracy of recognition.

The Conclusion and Prospect

This paper combines human pose estimation and action recognition to the future intelligent classroom, can help to improve classroom teaching efficiency and teaching quality, there is a broad application prospect in the future and the more important application value. Truly effective learning analysis not only at this point in the lesson, and before and after class teachers to carry out the analysis of the situation of learning an important opportunity. Pre-course reading speed, assignments of students and so on to determine the state of learning.

References

[3] Many learners gesture recognition based on depth image, such as computer science. Zhang Hongyu 2015.9.
[5] Introduction to psychology and teaching in the class of students In 1999, 6 Zhu Yanhua education research.
[7] Based on large data of students in the classroom teaching accurately focus on experimental research Yi-wen zheng in February 2016.