ISSN 1728-7901

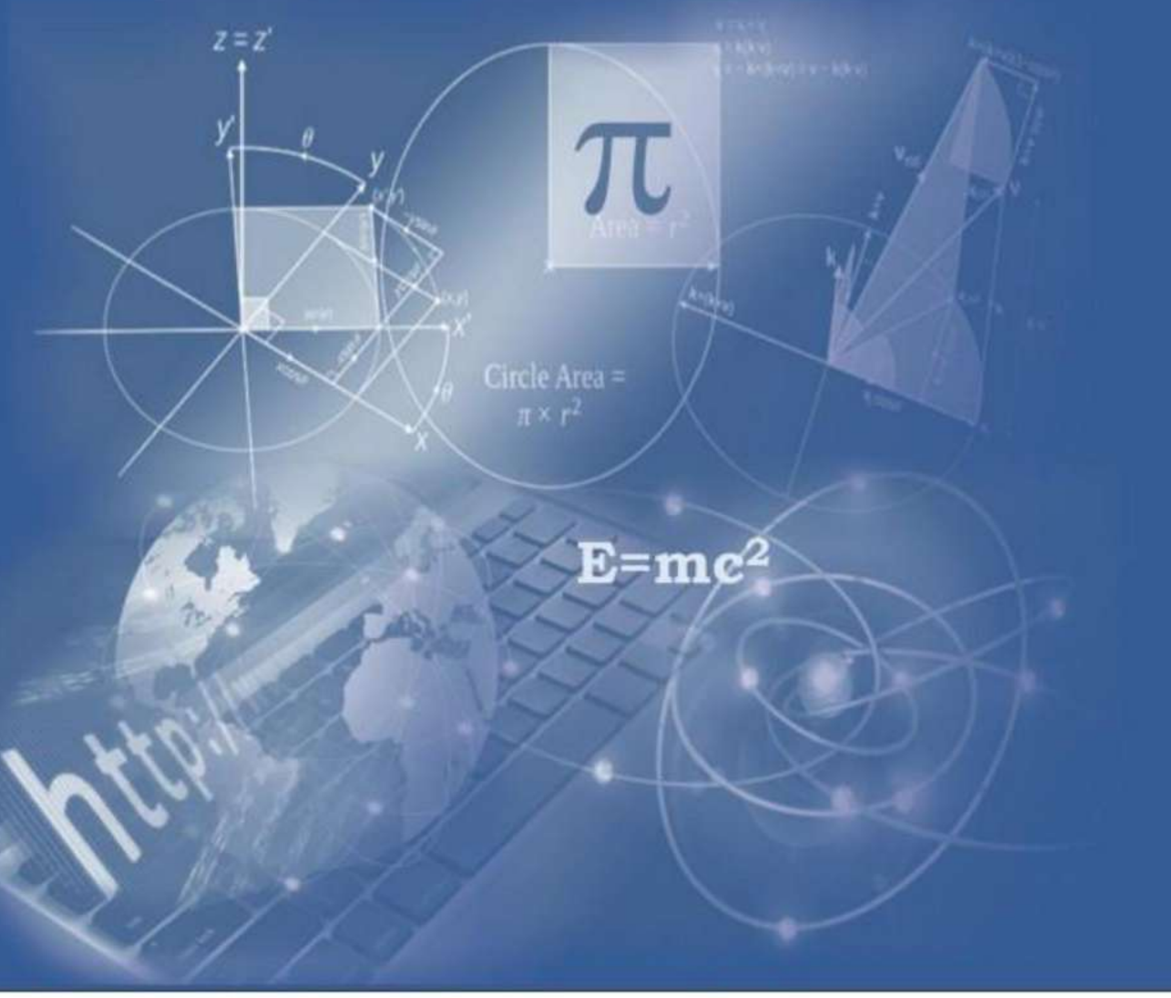
Абай атындағы Қазақ ұлттық педагогикалық университеті

Казахский национальный педагогический университет имени Абая

ХАБАРШЫ
ВЕСТНИК
BULLETIN

«Физика-математика ғылымдары» сериясы

серия «Физико-математические науки»



№ 4 (68)

2019

ISSN 1728-7901

Абай атындағы Қазақ ұлттық педагогикалық университеті
Казахский национальный педагогический университет имени Абая
Abai Kazakh National Pedagogical University

ХАБАРШЫ

«Физика-математика ғылымдары» сериясы
Серия «Физико-математические науки»
Series of Physics & Mathematical Sciences
№4(68)

Алматы, 2019

Казахский национальный педагогический университет имени Абая

ВЕСТНИК
Серия «Физико-математические науки»
№ 2 (66), 2019 г.

Зарегистрирован в Министерстве информации
Республики Казахстан,
№ 4824 - Ж - 15.03.2004
(периодичность – 4 номера в год)
Выходит с 2000 года

Подписано в печать 09.12.2019 г.
Формат 60х84 1/8. Об. 33,6 уч.-изд.л.
Тираж 300 экз. Заказ 252.

050010, г. Алматы, пр. Достык, 13,
Издательство «Ұлағат» КазНПУ им. Абая

ФИЗИКА. ФИЗИКАНЫ ОҚЫТУ ӘДІСТЕМЕСІ

ФИЗИКА. МЕТОДИКА ПРЕПОДАВАНИЯ ФИЗИКИ

ИНФОРМАТИКА. ИНФОРМАТИКАНЫ ОҚЫТУ ӘДІСТЕМЕСІ. БІЛІМ БЕРУДІ АҚПАРАТТАНДЫРУ ИНФОРМАТИКА, МЕТОДИКА ПРЕПОДАВАНИЯ ИНФОРМАТИКИ. ИНФОРМАТИЗАЦИЯ ОБРАЗОВАНИЯ

Пайдаланылған әдебиеттер тізімі:

1 Қазақстан Республикасындағы білім беруді дамытудың 2005-2010 жылдарға арналған мемлекеттік бағдарламасына түсінктеме.//Егеменді Қазақстан. А, №258, 16 қазан 2004 жыл, 5 б.

2 Роберт И.В. Новые информационные технологии в обучении: дидактические проблемы, перспективы использования. // Информатика и образование. М., 1991, №4, с. 18-25

3 Смағұлова Л.А. Орта мектептегі информатика курсында компьютерлік модельдеуді оқыту әдістемесін жетілдіру. Педагогика ғылымдарының кандитаты ғылыми дәрежесін алу үшін дайындалған диссертация. Алматы, 2002, 101 б.

4 Абдыкеримова Э.А. Компьютерлік модельдеудің негізінде мектеп физикасын оқытудың әдістемесі. Педагогика ғылымдарының кандитаты ғылыми дәрежесін алу үшін дайындалған автореферат. Алматы, 2004, 16.04.2004

5 Мәлібекова М.С. Жаңа ақпараттық технологияны математика және информатика пәндері байланысында қолданудың педагогикалық негіздері. Педагогика ғылымдарының кандитаты ғылыми дәрежесін алу үшін дайындалған диссертация. Алматы, 1999, 74 б.

МРНТИ 20.53.21, 20.47.23
УДК 004.932

APPLICATION OF THE MACHINE-LEARNING ALGORITHM FOR HUMAN EMOTIONS RECOGNITION

Zhaksylyk N.B. [1]*, Daribayev B.S.* [1]

[1]*Al-Farabi Kazakh National University, Almaty, Kazakhstan*

Abstract

Modern methods of deep learning of neural networks are to find the minimum of some continuous error function. Currently, various optimization algorithms are known that use different approaches to update model parameters. This article is devoted to the analysis of convolutional neural networks with a teacher, as well as the most common optimization methods for learning deep networks in order to recognize the human emotion in the picture. Learning with a teacher is one of the ways of machine learning, during which the test system is forcibly trained with the help of “stimulus-reaction” examples. During the analysis, activation functions such as ReLU, sigmoid were used. And also, categorical_crossentropy was chosen to find errors in the network. And they minimized the error by updating weights to correctly output the response using the Adam optimizer. Corrected a network error using the Dropout regulator.

Keywords: Multilayer perceptron, neural network, activation function, weighting factors, tensors, convolutional neural network, gradient descent.

Аннотация

Н. Б. Жақсылық [1]*, Б.С. Дарибаев* [1]

ПРИМЕНЕНИЕ АЛГОРИТМА МАШИННОГО ОБУЧЕНИЯ ДЛЯ РАСПОЗНАВАНИЯ ЭМОЦИЙ ЧЕЛОВЕКА

[1]*Казахский Национальный Университет им. аль-Фараби, г. Алматы, Казахстан*

Современные методы глубокого обучения нейронных сетей заключаются в нахождении минимума некоторой непрерывной функции ошибки. В настоящее время известны различные алгоритмы оптимизации, которые используют разные подходы для обновления параметров модели. Данная статья посвящена анализу сверточных нейронных сетей с учителем, а также, наиболее распространенных методов оптимизации для обученя глубоких сетей чтобы распознать эмоцию человека на картинке. Обучение с учителем — один из способов машинного обучения, в ходе которого испытуемая система принудительно обучается с помощью примеров «стимул-реакция». В процессе анализа были использованы активационные функции такие как ReLU, sigmoid. А также, для нахождения ошибки в сети выбрали categorical_crossentropy. И минимизировали ошибку обновляя весовые коэффициенты для корректного вывода ответа с помощью оптимизатора Adam. Исправляли ошибку сети используя регулизатор Dropout.

Ключевые слова: Многослойный персептрон, нейронная сеть, активационная функция, весовые коэффициенты, тензоры, сверточная нейронная сеть, градиентный спуск.

Аңдатпа

Н. Б. Жақсылық [1], Б.С. Дарибаев [1]

АДАМ ЭМОЦИОНАЛАРЫН АНЫҚТАУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ АЛГОРИТІМІН ПАЙДАЛАНУ

[1]аль-Фараби атындағы Қазақ Ұлттық университеті, Алматы қ., Қазақстан

Нейрондық желілерді терең оқытудың заманауи әдістері кейбір қателіктер функциясының минимумын табу болып табылады. Қазіргі уақытта модель параметрлерін жаңарту үшін әртүрлі тәсілдерді қолданатын әртүрлі оңтайландыру алгоритмдері белгілі. Бұл мақала мұғалімнің көмегімен конвульсиялық нейрондық желілерді, сонымен қатар суреттегі адамның эмоциясын тану үшін терең желілерді оқытудың оңтайландыру әдістерін талдауға арналған. Мұғаліммен оқыту - машиналық оқытудың бір әдісі болып табылады, оның барысында тест жүйесі «ынталандыру-реакция» мысалдары негізінде мәжбүрлі түрде оқытылады. Талдау кезінде ReLU, sigmoid сияқты белсендіру функциялары қолданылды. Сондай-ақ, желідегі қателерді табу үшін categorical_crossentropy таңдалды. Олар Adam оңтайландырушының көмегімен жауапты дұрыс шығару үшін салмақты жаңарту арқылы қатені азайттық. Dropout реттегішінің көмегімен желілік қатені дұрыстадық.

Түйін сөздер: Көп қабатты перцептрон, нейрондық желі, белсендіру функциясы, салмақ факторлары, тензорлар, конвульсиялық нейрондық желі, градиенттің түсуі.

Introduction

People communicating with each other constantly analyze any human manifestations. It is no secret that one of the important aspects of the analysis is human emotions. Therefore, the creation of modern technology of machine systems is relevant application of methods for automatic recognition of emotions.

One of the common ways of recognizing human emotions by another person is the analysis of visual information. Based on this, it can be understood that the automation of this process should be obvious based on the use of computer vision methods, which is implemented using modern technology of neural networks.

Computer vision is a branch of science where theory and methods are studied, as well as algorithms for analyzing images of objects [1]. This technology is the creation of machines that can detect, track and classify objects. As a scientific discipline, computer vision refers to the theory and technology of creating artificial systems that receive information from images.

Automatic recognition of emotions is quite widespread. For example, recently, Kia Motors introduced [2] a system that can assess the mood of a person in a car and adjust the atmosphere in the cabin for him.

Such artificial systems are well-implemented using convolutional neural networks. At present, the use of convolutional neural networks allows one to recognize human emotion with various methods of data analysis. For example, one can recognize a person’s emotion with the help of an audio recording implemented by Reza Chu [3]. Using this technology, we implemented a network that can recognize the emotion in the picture.

In order to recognize human emotion, a convolutional neural network (CNN) [4] was used, which was trained based on Kaggle [5] data, where gray images with 1 channel are stored in 48x48 pixels format. For 2 reasons, we used this particular method and not a simple multilayer perceptron [6]:

• The percentage of accuracy on the verification data is greater. Due to this, the computer determines the emotion of a person in our case better in our cases.

• The concept of total weights.

Let's look at it better to understand the above-mentioned pluses of the CNN. First, let's see how version 2 of the usual multilayer perceptron works.

A multilayer perceptron consists of an input, hidden (there may be several of them, depending on the task), output layer (Figure 1).

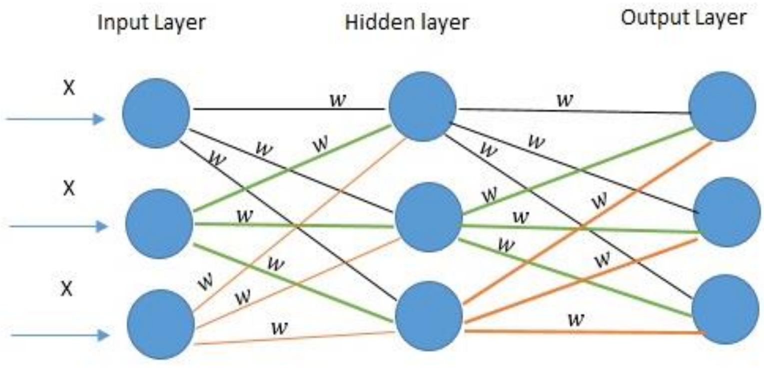


Figure 1. The structure of neural networks

As you can see, each node (neuron) in each layer is interconnected. The binding strands of neurons are called synapses, which have their own weights (weighting factors). In machine learning, weights are usually marked with the letter "w", and "X" is the input data that is transmitted to the first layer. To get a response from the output layer, in each layer, all neurons must perform several operations and transfer data to the last layer.

Accordingly, each neuron is 1 pixel of the picture. If we multiply the width of 48px by the height of 48px, then we get the value 2304. Our input layer would have 2304 neurons and since each neuron in each layer is interconnected, we would have many weighting factors. And this is not beneficial for us since it would require a large amount of resources. Conversely, convolutional neural networks are limited by the number of weights. And this has become the main reason for the success of the CNN in the recognition of large objects such as pictures, audio, video, etc. Since our emotions are tensors (48x48 matrix), we used this method to train our model.

To date, the best results in image recognition are obtained with their help. On average, the recognition accuracy of such networks exceeds conventional perceptron by 10-15%. CNN is the core technology of Deep Learning.

The main difference between a fully connected layer and a convolutional layer is the following: multilayer perceptron layers study global patterns in the space of input features (for example, in the case of emotions from the Kaggle set, these patterns involve all pixels), while convolutional layers study local patterns (Figure 2).

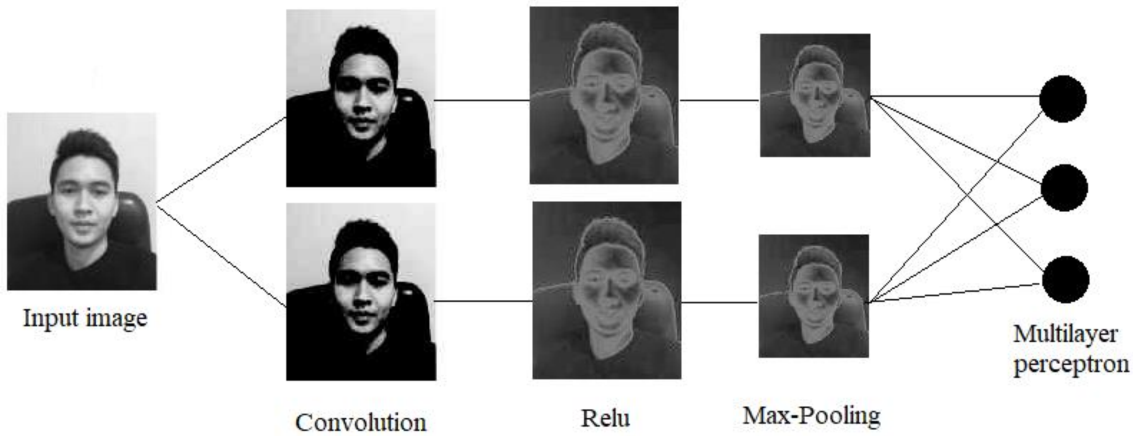


Figure 2. Convolution operation

Input layer

The input data are gray images of the type JPEG, size 48x48 pixels. If the size is too large, the computational complexity will increase, respectively, the restrictions on the response speed will be violated, the determination of the size in this problem is solved by the selection method. If you select a size too small, then the network will not be able to identify key signs of faces.

If you look at Figure 2, you can see that a picture with human emotions is a picture of only 1 channel (48x48x1). The input layer takes into account the two-dimensional topology of the images and consists of one map (matrices), the map can be one, if the image is presented in shades of gray, otherwise there are 3, where each map corresponds to an image with a specific channel (red, blue and green).

The input data of each specific pixel value is normalized to a range from 0 to 1, according to the formula:

$$f(p, \min, \max) = \frac{p - \min}{\max - \min}$$

where,

$$f - normalization\ function.$$
$$p - specific\ color\ value\ in\ pixels\ from\ 0\ to\ 255.$$
$$\min - minimum\ pixel\ value - 0.$$
$$\max - maximum\ pixel\ value - 255.$$

Convolutional layer

Convolution is applied to three-dimensional tensors, called feature maps, with two spatial axes (height and width), as well as with the depth axis (or the axis of the channels). For black and white images, as in the Kaggle dataset, the depth axis has a dimension of 1 (shades of gray). The collapse operation extracts patterns from its input feature map and applies the same transformations to all patterns, producing an output feature map. The size of the output card can be calculated using this formula:

$$(w,h) = mW - kW + 1, mH - kH + 1$$

where,

$$(w,h) - calculated\ convolution\ card\ size.$$
$$mW - width\ of\ previous\ map.$$
$$\mathrm{h}W - height\ of\ previous\ map.$$
$$kW - core\ width.$$
$$kH - core\ height.$$

The core is a filter or a window that slides over the entire area of the previous map and finds certain signs of objects. Since we trained the network on many faces in order to recognize human emotion, one of the cores could give the greatest signal in the process of training in the mouth, eyebrow, eye and nose (Figure 3).

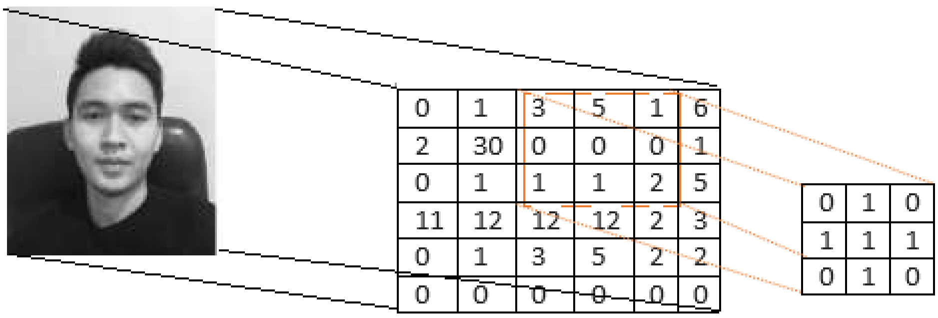


Figure 3. The convolution operation and obtaining the value of the convolution card

The kernel glides over the previous map and performs the convolution operation and transfers the values to the activation function, then, the characteristic maps are created (new matrix), the formula:

$$(f * g)[m,n] = \sum_{k,l} f[m-k, n-l] * g[k,l]$$

where,

$$f - source\ image\ matrix.$$
$$g - convolution\ core.$$

Relu was chosen as the activation function, since it does not lead to problems with attenuation or increasing gradients. And also, for the hidden layers of the perceptron, we used Relu, and the output was Sigmoid.

$$f(s) = \max(0, s) - activation\ function\ Relu.$$

The operation of this function is simple, as you can see this, the function displays 0 if the value passed to the function is $s < 0$, and if $s \geq 0$ returns the same value.

Having performed the activation function, we transfer the resulting matrix to the subsample layer (Figure 4).

The purpose of the layer is to reduce the dimension of the maps of the previous layer. If some signs were already detected in the previous convolution operation, then such a detailed image is no longer needed for further processing, and it is compressed to a less detailed one. In addition, filtering already unnecessary parts helps not to retrain.

During scanning by the core of the subsample layer (filter) of the map of the previous layer, the scanning core does not intersect, unlike the convolutional layer. Usually, each card has a 2x2 core, which allows you to reduce the previous card convolution layer by 2 times.

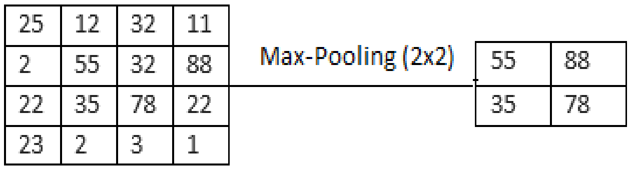


Figure 4. Layer subsampling

In our problem, there will be 3 convolutional layers of 1 subsample layer for each and a multilayer perceptron with 1 hidden layer.

As you can see after reducing the size, we transfer the reduced image size to the input of the multilayer perceptron without losing important information.

After receiving the vector data, we carry out the training. The essence of training is to reduce the function of loss. This is implemented using the backpropagation method. First, we find the exit error by the formula:

$$E = (d - y)^2$$

where,

$$d - expected\ result.$$
$$y - output.$$

The output is calculated using the Sigmoid activation function. The sigmoid is expressed by the formula:

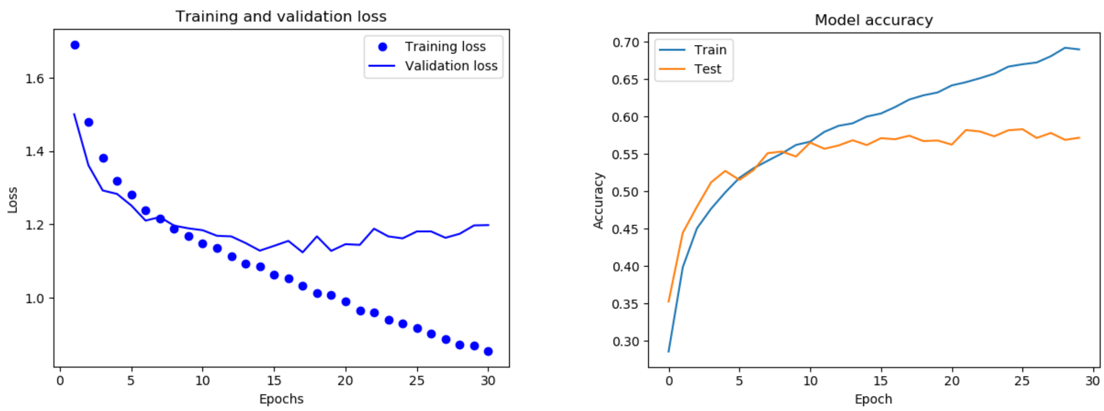
$$f(s) = \frac{1}{1 + e^{-s}}$$

where,

$$S = \sum_{i=1}^{n} x_i * w_i$$
$$n - number\ of\ neurons.$$

As we said above, in order for the network to work correctly, we must reduce losses. This is implemented using the gradient descent method [7]. For this we will use the Adam optimizer [8, 9]

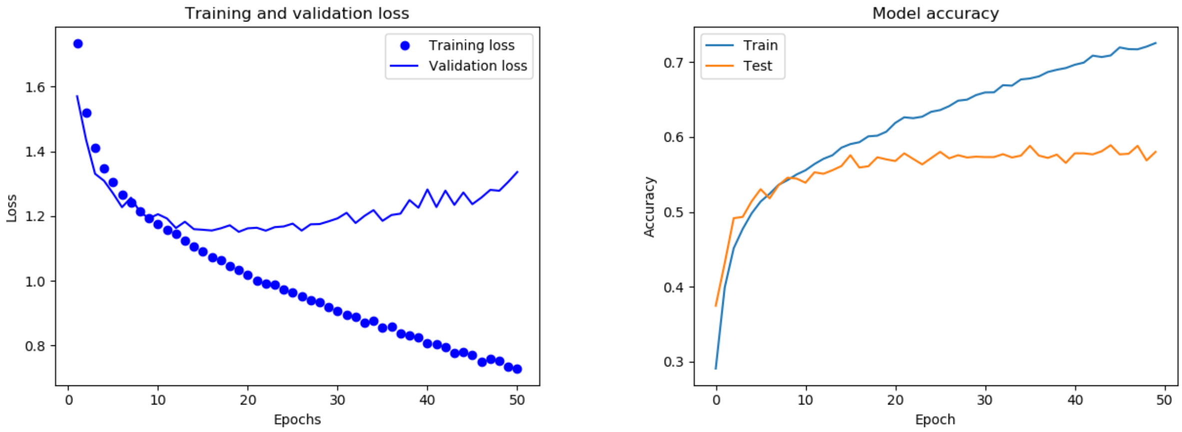
During the first experiment, we obtained results of loss and network accuracy (Figure 5a, Figure 5b):



a) b)

Figure 5.
a) Losses on training and validation data in the 30th epochs.
b). Accuracy on training and validation data in the 30th epochs

Increasing the number of epochs from 30 to 50, we got the following results (Figure 6a, Figure 6 b):



a) b)

Figure 6.
a) Losses on training and validation data in the 50th epochs.
b). Accuracy on training and validation data in the 50th epochs

Entrance training model we are faced with retraining. This term means that our network recognizes emotion on the training data well, however, the accuracy on the validation data did not rise. Due to this, the method of thinning (Dropout) was used.

Retraining is a problem for many neural networks. Due to fact that the network begins to memorize training data, our network will begin to learn excessively. Any method that we use to prevent retraining is called the regularization method.

One of these regularization methods is loss (thinning), which is used in deep networks by including drop-down layers. This layer does not contain neurons; accordingly, it does not calculate anything. Instead, it temporarily disconnects some neural from the previous layer. Such a layer will be temporarily active in the learning process. When we use the network for prediction, the drop-down layer does not work.

Conclusion

During the study, we achieved 60% accuracy in recognizing human emotions on validating data. This is not a bad result for not more data. By further increasing our data, we will improve prediction accuracy.

The given approach to automatic recognition of emotions can be effectively applied in various intelligent human-machine systems.

References:

1 Klette R., Komp'yuternoye zreniye. Teoriya i algoritmy – 2019

2 Raspoznavaniye emotsiy, podstroit salon bespilotnogo avtomobilya pod nastroyeniye passazhira [An electronic resource]. – 2019. – URL: https://nplus1.ru/news/2019/01/07/kia-read (Date of the application: 13.06.2019)

3 Speech Emotion Recognition with Convolutional Neural Network [An electronic resource]. – 2019. – URL: https://towardsdatascience.com/speech-emotion-recognition-with-convolution-neural-network-1e6bb7130ce3 (Date of the application: 18.09.2019)

4 Chollet Francois, Glubokoye obucheniye na Python. — SPb.: Piter, 2018. — p. 400

5 Data sources [An electronic resource]. – 2018. – URL: https://www.kaggle.com/deadskull7/fer2013 (Data of the application: 11.05.2019)

6 Tariq Rashid, Make your own neural network – 2018

7 Glassner E., Glubokoye obucheniye bez matematiki. T. 1: Osnovy / per. s ang. V. A. Yarotskogo. – M.: DMK Press, 2019. – p. 578

8 Гусманова Ф.Р., Абдулкаримова Г.А. Имитациялық модельдеудің құралдарына шолу // Вест. КазНПУ, Физико-математические науки, 2 (66), 2019, С. 234-240

9 Metody optimizatsii neyronnykh setey [An electronic resource]. – 2017. – URL: https://habr.com/ru/post/318970/ (Data of the application: 07.10.2018)