МРНТИ 20.53.21, 20.47.23
УДК 519.683

RECOGNITION OF THE TEXT BY MEANS OF DEEP LEARNING

Nurmukhanov T.A. [1], Daribayev B.S. [1]

[1]AL-Farabi Kazakh National University, Almaty, Kazakhstan

Abstract

Using neural networks, various variations of the classification of objects can be performed. Neural networks are used in many areas of recognition. A big area in this area is text recognition. The paper considers the optimal way to build a network for text recognition, the use of optimal methods for activation functions, and optimizers. Also, the article checked the correctness of text recognition with different optimization methods.

This article is devoted to the analysis of convolutional neural networks. In the article, a convolutional neural network model will be trained with a teacher. Teaching with a teacher is a type of training for neural networks in which you provide the input data and the desired result, that is, the student looking at the input data will understand that you need to strive for the result that was provided to him.

Keywords: neural network, deep learning, machine learning.

Аннотация

Т.А. Нурмуханов[1], Б.С. Дарибаев[1]
[1]Казахский Национальный Университет им. аль-Фараби, г. Алматы, Казахстан

РАСПОЗНАВАНИЕ ТЕКСТА ПРИ ПОМОЩИ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ

При помощи нейронных сетей можно выполнить разные вариации классификации объектов. Нейронные сети применяется во многих областях распознавания. Большой областью в этой сфере является распознавание текста. В работе рассматривается оптимальный способ построение сети для распознавания текста, применению оптимальных методов для активационных функции, и оптимизаторов. Также в статье проверили корректность распознавания текста при разных методах оптимизации.

Данная статья посвещена анализу сверточных нейронных сетей. В статье модель сверточных нейронных сетей будет обучаться с учителем. Обучение с учителем – это тип обучения нейронных сетей в котором вы предоставляете входные данные и желаемый результат, то есть ученик посмотрев на входные данные поймет, что нужно стремиться к тому результату который ему предоставили.

Ключевые слова: нейронные сети, глубокое обучение, машинное обучение.

Аңдатпа

Т.А. Нурмуханов[1], Б.С. Дарибаев[1]
[1]аль-Фараби атындағы Қазақ Ұлттық Университеті, Алматы қ., Қазақстан

ТЕРЕҢ ОҚЫТУ ӘДІСТЕРІ АРҚЫЛЫ МӘТІНДІ ТАНУ

Нейрондық желілерді қолдана отырып, объектілерді жіктеудің әртүрлі вариацияларын орындауға болады. Нейрондық желілер танудың көптеген саласында қолданылады. Бұл саладағы үлкен бағыт - мәтінді тану. Мақалада мәтінді тану үшін желіні құрудың оңтайлы тәсілі, активтендіру функциялары мен оптимизаторлар үшін оңтайлы әдістерді қолдану қарастырылған. Сонымен қатар, мақала әр түрлі оңтайландыру әдістерімен мәтінді танудың дұрыстығын тексерді. Бұл мақала консолюциялық нейрондық желілерді талдауға арналған. Мақалада мұғалімнің көмегімен нейрондық жүйесінің үйірткі моделі оқытылады. Мұғаліммен сабақ беру - бұл нейрондық желілерге арналған оқыту түрі, онда сіз енгізілген мәліметтерді және қалаған нәтижені бересіз, яғни кіріс деректерін қарап жатқан студент сіз оған берілген нәтижеге ұмтылуыңыз керек екенін түсінеді.

Түйін сөздер: нейронды желілер, терең оқыту, машиналық оқыту.

Introduction

Today, machine learning is actively used to recognize images and objects, and many search engines are built on its basis. One of the most common pattern recognition tasks is text recognition. Despite the large number of different programs designed for text recognition, the relevance of developing new software tools does not decrease. Text is an excellent means of communication and documentation. It remains the most effective and easiest way to express a person’s thoughts. During the rise of the computer era, when using scanned text in computers, it was less convenient to use these documents on an industrial scale, but with the development of information technologies and devices, as well as the beginning of the digitalization of documents, it became an urgent task to convert scanned text to computer-readable data task for today.

The active development of machine learning has led to the proliferation of artificial neural networks. One of the most effective types of artificial neural networks for text recognition is the convolutional neural network.

The operation of a convolutional neural network is usually interpreted as a transition from specific features of the image to more abstract details, and then to even more abstract details, up to highlighting concepts of a high level. Despite their large size, these networks have a small number of configurable parameters. This article will examine convolutional neural networks for text recognition.

Research in the field of text recognition has been going on for a very long time. For example, in 1993, the text recognition technology of the Russian company ABBYY was released. On its basis, a number of corporate solutions and programs for mass users have been created. In particular, ABBYY FineReader, a program for recognizing texts, applications for recognizing text information from mobile devices, ABBYY FlexiCapture [1], a system for streaming input of documents and data.

Convolutional neural networks are used to implement such tasks. Using this technology allows you to recognize text with very high accuracy. For example, network can recognize handwritten text that Priya Dwivedi [2] implemented in English. The article will consider the main stages in the construction of a neural network for text recognition.

Introduction

To test the developed system, a database was used containing various styles of English letters and their corresponding labels. For the training of the network, 370000 images were used, of which 279337 images were used as a training sample and 93113 images as a test one. The size of each image is 28x28 pixels. Each pixel is encoded by a number in the interval [0; 1], where 0 corresponds to black and white to 1.

Figure 1. Picture of educating for letter u

Image pre-processing is necessary to achieve maximum system recognition accuracy. Thus, the resulting neural network will be invariant to minor distortions such as noise, rotation and scaling.

Simple image distortions such as shift, rotation and angular displacement can be eliminated by applying simple affine transformations. Also, to improve the quality of the neural network, the images supplied to the input were additionally averaged.

This approach allows you to get a network, the training of which was carried out using a different set of images for each era.

Description of neural networks

A convolutional neural network model containing 3 layers was implemented and used in the work. Network training took place over 20 eras. In this article, a convolutional network was chosen for 2 reasons:

• Using common weights for each layer reduces the number of customizable parameters.

• Greater accuracy on validation data.

For clarity, it is necessary to compare convolutional neural networks with multilayer perceptrons.

A multilayer perceptron consists of an input, hidden (their number varies from 1 to several tens), an output layer (Fig. 2).

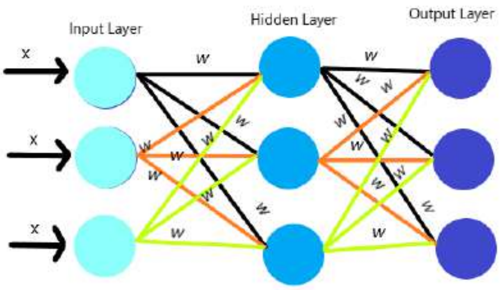


Figure 2. Structure of neural network

The connecting strands of neurons are called synapses, which have their own weight coefficients. In the field of machine learning, weight coefficients are marked with the letter “w”, and “x” is the input data that is transmitted to the first layer. In order to transmit data for the input layer, in each previous layer, 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 28px by the height of 28px, we get the value 784. The input layer would have 784 neurons and since each neuron in each layer would be connected, the naya would have many weight coefficients. This is not profitable since it would require a large amount of resources.

Convolutional neural networks are limited by the number of weight coefficients. And this has become the main reason for the success of convolutional neural networks in the recognition of pictures, audio, video, etc. Since the letters in the network are a 28x28 matrix (tensors), this method was used for training in this model. The main difference between a fully connected layer and a convolutional layer is that convolutional neural networks study local patterns in the space of input features, while multilayer perceptrons study global patterns.

The picture with letters is a picture of only one channel (28x28x1). The input layer reads a two-dimensional image topology and consists of one matrix (map), there can be one map, if the image is presented in shades of gray, otherwise there are 3 of them, where each map corresponds to an image with a specific channel (red, green, blue) .

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

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

where,

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

Convolution is applied to three-dimensional tensors, called feature maps, with a depth axis (channel axis), and with two space axes (height and width). For black and white images, the depth axis has one dimension (shades of gray). The folding 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 map can be calculated using the formula:

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

where,

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

The core is a filter that glides over the entire area of the previous map and finds certain signs of objects. Since the network was trained on many letters in order to recognize the text, then one core in the learning process could give the largest signal in a certain area.

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

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

where,

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

Methods

In this article, three activation functions Relu, Elu, Tanh were compared and used. For hidden layers of the perceptron, these three functions were also used, and the Sigmoid function was used on the output layer. Having performed the activation function, the resulting matrix is transferred to the subsample layer. 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. By the way, filtering out already unnecessary details helps not to retrain.

During the scanning process by the filter of the card of the previous layer, the scanning cores do not intersect, unlike the convolution layer. Each card has a core size of 2x2, which allows you to reduce the previous card convolution layer by 2 times.

After reducing the size, the reduced image size is transmitted without losing important data to the input of the multilayer perceptron.

After receiving the vector data, training is performed. The task of training is to reduce the function of loss. This is implemented using the backpropagation method.

The first step is to find errors by the formula:

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

where,

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

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

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

where,

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

Relu

$$f(s) = \max(0, s)$$

Relu activation function.

The function displays 0 if the value passed to the function is s <0, and if s≥0 returns the same value.

Elu

$$f(s) = (s, alpha = 1.0)$$

Elu activation function.

The function displays s if the value passed to the function is s> 0, and if s <0 returns: $alpha * e^s - 1$.

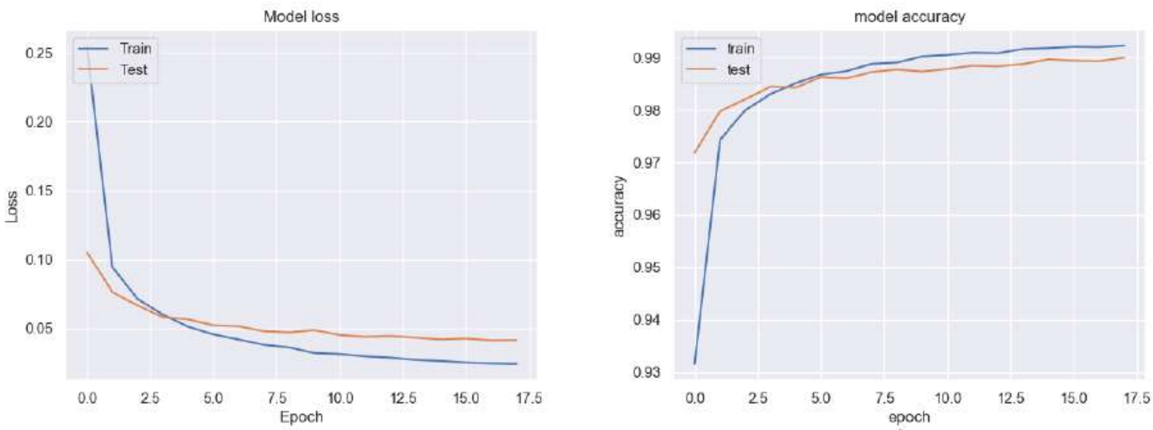
Tanh

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

Tanh activation function.

Results

The most inefficient network in this article is the network where the Tanh activation function was used. The accuracy of this network on verification data is 99.00%, and on training data 99.23%. The loss of this network is 0.023.



a) b)

Figure 3. a) Losses on training and verification data. b) Accuracy on training and verification data.

The average network in this article is the network where the Elu activation function was used. The accuracy of this network on verification data is 99.29%, and on training data 99.50%. The loss of this network is 0.0147

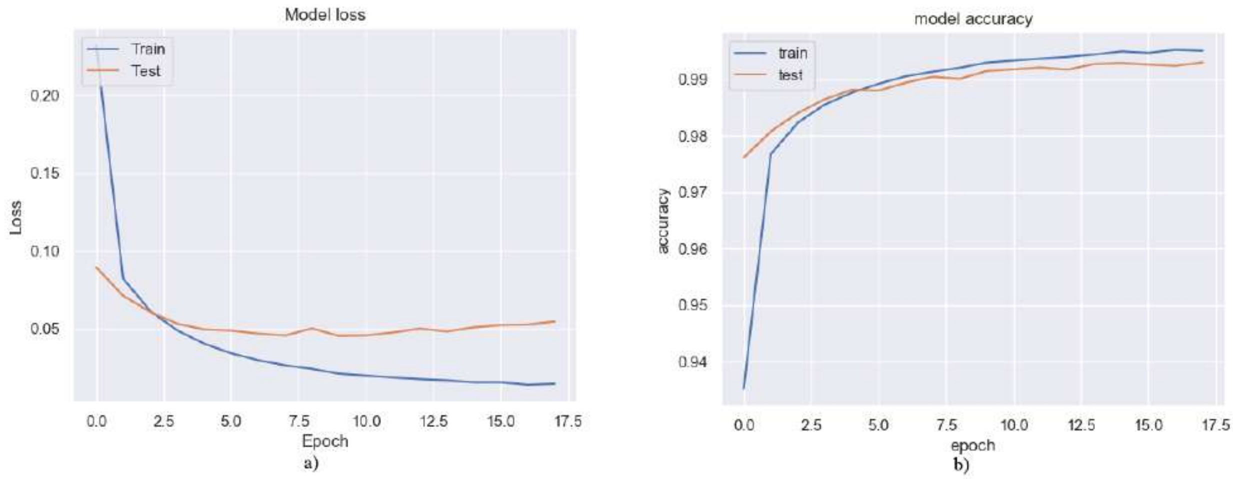


Figure 4. a) Losses on training and verification data. b) Accuracy on training and verification data.

The most effective network in this article is the network where the Relu activation function was used. The accuracy of this network on verification data is 99.38%, and on training data 99.63%. The loss of this network is 0.0107.

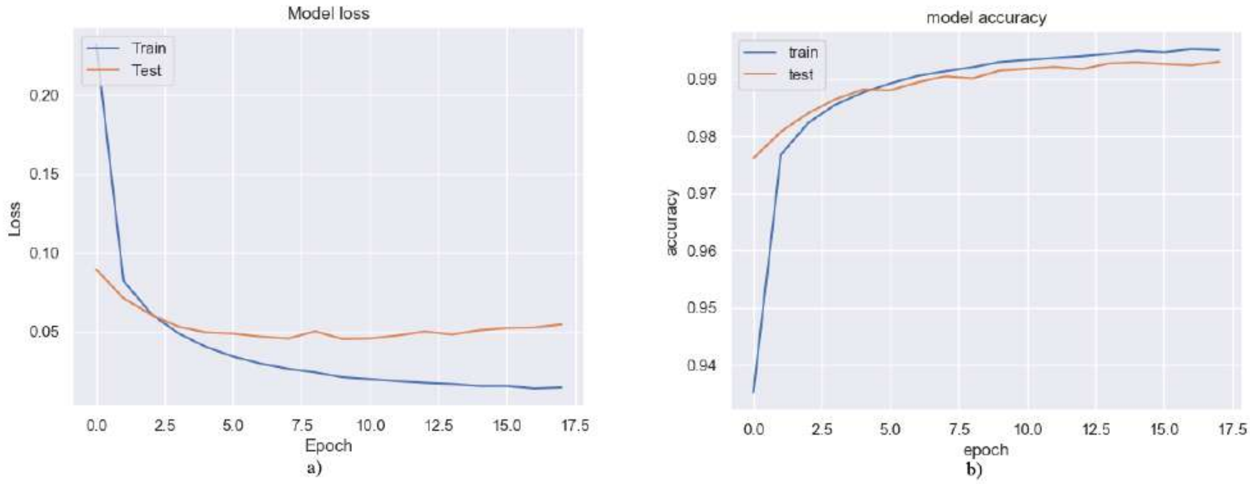


Figure 5. a) Losses on training and verification data. b) Accuracy on training and verification data.

Discussion

During training, the network faced retraining. This means that the network ceases to correctly recognize verification data due to incorrect memorization of patterns.

We solved the problem with retraining the network by selecting the optimal number of layers and eras.

Retraining is a common problem for neural networks. Due to the fact that the network begins to memorize training data, the network will begin to learn excessively. The method that is used to prevent overfitting is called the regulatory method.

Conclusion

During the study, the recognition accuracy of letters was achieved over 99% on the validating data. Three types of activation functions are compared and their accuracy is shown.

The above approach to automatic recognition of letters can be effectively applied in various text recognition systems.

References:

1 Распознавание эмоций подстроит салон беспилотного автомобиля под настроение пассажира [An electronic resource]. – 2019. – URL: https://ru.wikipedia.org/wiki/Оптическое_распознавание_символов (Date of the application: 16.08.2019)

2 Handwriting recognition using Tensorflow and Keras [An electronic resource]. – 2018. – URL: https://towardsdatascience.com/handwriting-recognition-using-tensorflow-and-keras-819b36148fe5 (Date of the application: 25.01.2018)
3 Build a Handwritten Text Recognition System using TensorFlow [An electronic resource]. – 2019. – URL: https:// towardsdatascience.com/build-a-handwritten-text-recognition-system-using-tensorflow-2326a3487cd5 (Date of the application: 15.06.2018)
4 Data sources [An electronic resource]. – 2018. – URL: https:// www.kaggle.com/sachinpatel21/az-handwritten-alphabets-in-csv-format (Data of the application: 16.02.2018)
5 Glassner E., Glubokoye obucheniye bez matematiki. T. 1: Osnovy / per. s ang. V. A. Yarotskogo. – M.: DMK Press, 2019. – p. 578
6 Методы оптимизации нейронных сетей [An electronic resource]. – 2017. – URL: https://habr.com/ru/post /318970/ (Data of the application: 07.10.2018)

МРНТИ 27.43.51
УДК 519.688

Е.Г. Неверова

Университет НАРХОЗ, г. Алматы, Казахстан

ИССЛЕДОВАНИЕ ДИНАМИКИ СПРОСА НА КРЕДИТОВАНИЕ ФИЗИЧЕСКИХ ЛИЦ С ПОМОЩЬЮ ИНСТРУМЕНТОВ ЯЗЫКА R

Аннотация

В статье проводится исследование оценки спроса физических лиц на услуги кредитования, предоставляемые банками второго уровня в Республике Казахстан. Источником информации послужил портал открытых данных Национального банка Республики Казахстан. Статистическим инструментом исследования выбран анализ временных рядов, выполненный средствами, предоставляемыми языком статистической обработки данных R. Выводы, сделанные на основании приведенных данных, доказывают, что среда программирования R имеет все возможности для быстрого, простого и наглядного анализа временных рядов при использовании специализированных пакетов, в том числе пакета forecast. Прогнозирование временных рядов проведено методом экспоненциального сглаживания. Тренды, обнаруженные в рядах для различных видов потребительского кредитования, показывают общую тенденцию к нестабильности и некоторому снижению спроса на данный вид услуг.

Ключевые слова: потребительское кредитование, ипотечное кредитование, прогнозирование, временные ряды, пакет forecast, язык R.

Аңдатпа

Е.Г. Неверова
NARHOZ университеті, Алматы қ., Қазақстан

R ТІЛДІК ҚҰРАЛДАРДЫ ҚОЛДАНА ОТЫРЫП ЖЕКЕ ТҰЛҒАЛАРДЫ НЕСИЕЛЕУГЕ СҰРАНЫС ДИНАМИКАСЫН ЗЕРТТЕУ

Мақалада жеке тұлғалардың Қазақстан Республикасындағы екінші деңгейлі банктер ұсынатын несиелік қызметтерге деген сұранысы зерттелген. Ақпарат көзі - Қазақстан Республикасы Ұлттық Банкінің ашық деректер порталы. Статистикалық зерттеу құралы статистикалық мәліметтерді өңдеудің тілінде берілген құралдармен орындалған уақытты талдауды таңдайды. Ұсынылған мәліметтерге негізделген тұжырымдамалар R бағдарламалау ортасы мамандандырылған пакеттерді пайдалану кезінде уақыт қатарларын жылдам, қарапайым және анық талдауға, болжам пакетін қоса алғанда. Уақыт қатарларын болжау экспоненциалды тегістеу әдісімен жүзеге асырылады. Тұтынушылық несиелеудің әртүрлі түрлерінің қатарында қалыптасқан тенденциялар тұрақсыздықтың жалпы тенденциясын және осы қызмет түріне сұраныстың аздап төмендеуін көрсетеді.

Түйін сөздер: тұтынушылық несиелеу, ипотекалық несиелеу, болжау, уақыт сериясы, forecast пакеті, R тілі.

Abstract

RESEARCH OF DYNAMICS OF DEMAND FOR LOANING OF INDIVIDUALS BY USING LANGUAGE INSTRUMENTS R

Neverova Ye.G.
NARHOZ University, Almaty, Kazakhstan

The article studies the assessment of the demand of individuals for lending services provided by second-tier banks in the Republic of Kazakhstan. The information source was the open data portal of the National Bank of the Republic of Kazakhstan. The statistical research tool selected time series analysis performed by the tools provided by the statistical