Proceedings of the
35th International Conference on Information Technologies (InfoTech-2021)
IEEE Conference, Rec. # 52438, 16-17 September 2021, Bulgaria

RECOGNITION OF HANDWRITTEN NUMBERS WITH DATA SETS

M. Fatih Cambek[1] and Pinar Kirci[2]

[1]Department of Computer Engineering, Faculty of Engineering, Istanbul University-Cerrahpasa, Istanbul
[2]Department of Computer Engineering, Faculty of Engineering, Bursa Uludag University, Bursa
pinarkirci@uludag.edu.tr
Turkey

Abstract: Today, transactions are still carried out with people in the transfer of documents and information, and as a result of these transactions, less favorable situations in terms of cost, time and resources are obtained. In the project, with the algorithm developed over the data sets, the possible writing forms of the numbers will be taught to the machine and it will be ensured that machine can distinguish the handwritten digit combinations they will encounter. With the presented machine learning project, it is aimed to define numbers in handwritten documents. In this way, it will reduce the need for human labor. In this way, it is aimed to accelerate document processing and to increase user satisfaction.
Key words: Handwriting, numbers, python

1. INTRODUCTION

Handwriting recognition is the recognition of handwritten letters, numbers and symbols by computer systems. Although it is quite easy for humans, it is a very difficult problem to perceive lines and curves on a ground as letters and numbers automatically and as meaningful words in a more advanced stage [1].

Today, with the developing computer technology, documents have started to be created and filled in computer environment. In this way, many transactions have become much faster, and the desired information is reached more easily. However, many documents had been printed, filled and stored on paper for years. Also, most documents are still available only on paper. Therefore, it is very important to digitize these documents and to transfer the information to the computer environment. The method followed in transferring the printed documents to the computer environment

is to enter this information one by one on the computer. As can be guessed, this process takes a lot of time and is very error prone. For this reason, it is very important to automatically transfer the documents to the computer environment. Such a system includes parts such as extracting the document structure, text recognition and verification of results using linguistics. Among these, the most important task falls on writing recognition [2].

Current technology is still at a limited level in handwriting recognition. Handwriting recognition is still not a fully solved problem. The difficulty in handwriting recognition is due to the large number of different typefaces and the fact that the letters are linked to each other, as well as differing from person to person. The style of writing can vary depending on the situation and the pen or paper used. Letters can be of very different shapes and sizes, depending on the writing style and speed of the people [3].

Machine Learning is a method that makes inferences from existing data using mathematical and statistical methods and makes predictions about the unknown with these inferences [4].

Machine Learning has many uses in handwriting and signature recognition systems, biometric recognition systems (fingerprint-iris-palm etc.), selection of spam e-mails, financial transactions such as credit risk account, robotics and natural language processing [5].

Number recognition is an important part of handwriting identification. The accurate transfer of numbers from text to digital media is very important in today's technology. In the project, this process is aimed to be done in a fast and accurate manner.

2. WORKS ON RECOGNITION OF HANDWRITING

In recent years, great progress has been made in the work on text recognition, some programs that automatically recognize printed texts (printed product or typewritten product or computer written product) on a clean and readable surface have begun to enter our lives. These programs, with a high percentage of recognition, reduce costs and make life easier in many institutions and companies [6].

Handwriting recognition is the recognition of handwritten letters, numbers and symbols by computer systems. The fact that it differs from person to person, the letters in the word are intertwined, and even a single letter is expressed in dozens of forms makes it difficult to read and process handwriting in electronic environment. The fact that it is not limited to the Latin alphabet but with different directions, shapes and signs in different languages makes it very difficult to recognize and distinguish handwriting [1].

In [3], a handwritten word was considered and the conversion of the handwritten text into digital form was done by classification. Two basic approaches have been used to accomplish this task: directly classifying words and characterizing

segmentation. First, the Convolutional Neural Network (CNN) was used with various architectures to train a model that could classify words correctly. For the second, using Convolutional Long Short Term Memory networks (LSTM) to create bounding boxes for each character. Later, segmented characters passed to a CNN for classification and then each word is reconstructed according to the classification and segmentation results.

The handwritten address interpretation research began at CEDAR [7] in 1987, funded by the United States Postal Service (USPS). The goal was to automate postal sorting through a system that can read a handwritten street address and postcode and encode each envelope with the destination address for machine sorting.

This research ultimately led to the development and deployment of a system that automates mailing sorting through image analysis, number recognition, word recognition, mail directory search, and assigning a barcode that specifies the destination address. Since field testing began in 1996, Handwriting Address Interpretation System (HWAI) has been implemented in all USPS mail processing centers [2].

The aim of the study conducted at Bilkent University was to make handwritten documents available throughout the history. Although it is easier to recognize texts written using the Latin alphabet and using certain writing styles, the use of Far Eastern or Arabic alphabets makes recognition difficult. Signature recognition, recognition of equations and special symbols are examples of other systems similar and complementary to handwriting recognition systems [6].

3. STRUCTURE OF THE PROJECT

The human visual system is not affected by the differences in the size and direction of the letters, whereas in an automatic system, these can cause major problems. Most of the time, the letters express something only according to the meaning in the sentence, otherwise it is not possible to recognize. All of these make the handwriting recognition problem a difficult and not yet fully solved problem, and the advantages and conveniences of an automatic handwriting recognition system increase its appeal. Once handwriting recognition methods are up and running, this will also save us from using the keyboard and allow us to write and draw much more naturally [6].

Handwriting can take many different forms. You can use letters that are written quite properly and separately, which we can call hand print, as well as all letters can be connected to each other, or a mixture of the two can be used. While separated letters are much easier to recognize, recognition of compound letters is a more difficult problem to solve because it requires separation of letters [8].

The two graphs below show the training curves on the datasets we have received.

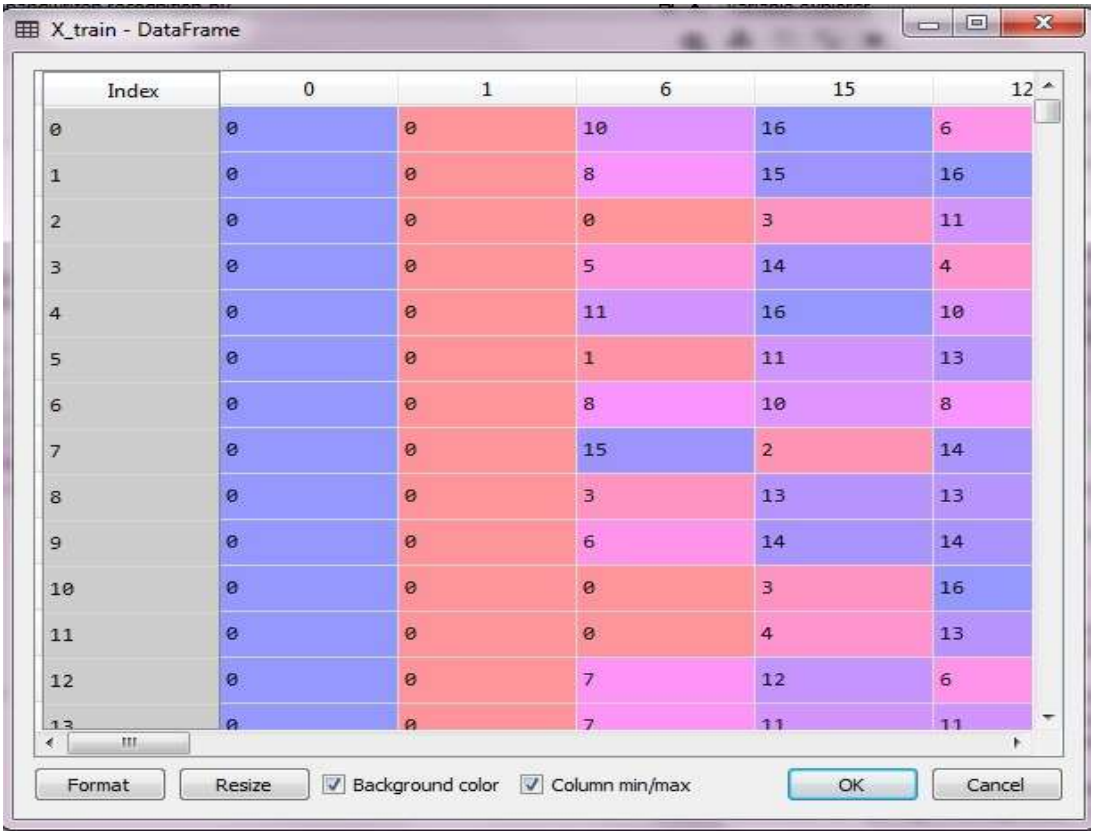

Fig. 1. The X_train table is created by ordering the data from the data sets by index numbers (3822).

The data in the datasets contain dozens of different spelling of the numbers as shown in figure 1 and figure 2.

Fig. 2. Data set content

Our data sets consist of lines as shown in figure 2. The data in these lines are divided into two parts. The first 64 data shows how the numbers entered are formed (indicated by a red line) and the number at the end (yellow dotted display) shows with which number these 64 data correspond to. The rightmost numbers (target) form our y train table.

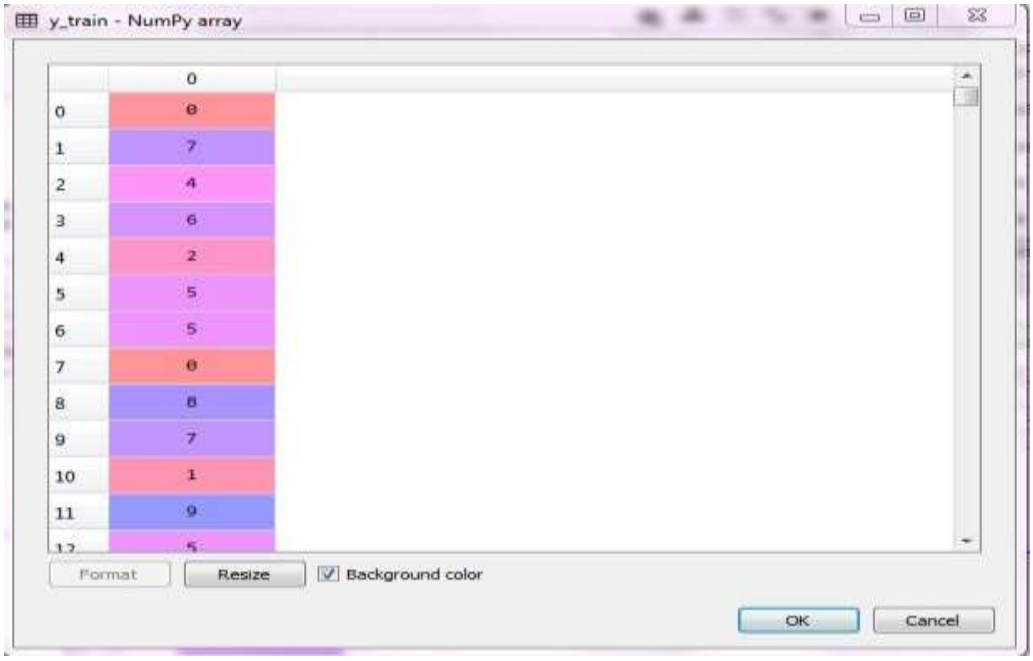

Fig. 3. Shows the results of numbers in data sets

In our Y_train table as given in figure 3, we have the table we call target, that is, match our results. It allows matching the numbers that come as a result of the rows we receive in the data sets with the results.

We plotted the first 50 data in our train data set in figure 4.

In figure 5, we see the first 50 data and estimated results in the test data set. Each number is estimated with its labels and the estimation results are written. The successful predictions are shown in green, the unsuccessful ones are shown in red.

Fig. 4. Top 50 Data Table of Train Data Set

Fig. 5. Test results

Python platform 3.6.4 version is used in programming. The training process was carried out by using pandas library methods in the python. In this study, dozens of

different spelling types of numbers were taught to the program with the first data sets. Then, the rate of matching the numbers in the data we tested with the ones entered was examined. In the light of the data obtained, the system has passed a certain success rate.

4. CONCLUSION

Writing is not only a learning space but also a communication tool, a tool for self-expression. Handwriting is the writing that comes out with handwriting tools such as pencil, chalk or charcoal. Since each person's handwriting is unique and different, it can be used to verify the author of a document. Although the meaning and importance of our handwriting in our daily life is gradually decreasing, it is still our unique feature, like our tone of voice, like our fingerprint.

Great emphasis is placed on the recognition of handwriting and its digital processing. Understanding the old documents correctly and recognizing the texts in the documents used today is an important issue focused on by different institutions.

REFERENCES

[1] About handwriting recognition, last access, 10-07-2020 (*available at*: https://msdn.microsoft.com/enus/library/windows/desktop/ms698602(v=vs.85).aspx)

[2] Bektas, B., Babur, S., Turhal, U., Köse, E., (2016). Optıcal character recognıtıon system vıa machıne learnıng. 5th International printing techologies symposium, Istanbul, November 2016.

[3] Balci, B., Saadati, D., Shiferaw, D. Handwritten Text Recognition using Deep Learning. Last access, 10-08-2020, (*available at*: http://cs231n.stanford.edu/reports/2017/pdfs/810.pdf)

[4] Makine ogrenmesi nedir?, last access, 11-08-2020, (*available at*: http://www.emrealadag.com/makine-ogrenmesi-nedir.html)

[5] Bilgin, M. Makine Öğrenmesi: Teorisi ve Algoritmaları. Editor, A. Yılmaz, ISBN 978-605-9594-25-7

[6] Duygulu, P. (2006). El yazısı tanıma. Bilişim Ansiklopedisi, papatya yayıncılık, (*available at*: http://www.cs.bilkent.edu.tr/~duygulu/)

[7] Cedar, last access, 11-07-2020, (*available at*: http://www.cedar.buffalo.edu/about.html)

[8] Koyun, A., Afsin, E. (2017). 2D Optical Character Recognition Based on Deep Learning. Türkiye bilişim vakfı bilgisayar bilimleri ve mühendisliği dergisi,Vol: 10, no:1