MBTI TYPE PREDICTION USING IMAGES FROM INSTAGRAM

MUSSAYEVA D.[1], KUANDYKOVA A.[2], TALASBEK A.[1], ORYNBEKOVA K.[1], KARIBOZ D.[1]

[1]*Suleyman Demirel University, 040900, Kaskelen, Kazakhstan*
[2]*JSC "Kompra", 010000, Nur-Sultan, Kazakhstan*

Abstract. *This article describes image based personality type prediction. With the help of the Instagram API, photos of respondents who previously passed the MBTI test were collected. As a result, 60 percent of accuracy was achieved. This work represents one of the solutions to determine what personality type a person belongs to without passing any tests, but simply uploading an image.*

Keywords: *Personality prediction, Image classification, CNN, Instagram API, MBTI.*

ИНСТАГРАМ ЖЕЛІСІНДЕГІ КЕСКІНДЕРДІ ҚОЛДАНУ АРҚЫЛЫ MBTI ТИПІН БОЛЖАУ

МУСАЕВА Д.[1], КУАНДЫКОВА А.[2], ТАЛАСБЕК А.[1], ОРЫНБЕКОВА К.[1], КАРИБОЗ Д.[1]

[1]*Сулейман Демирел атындағы университет, 040900, Қаскелең, Қазақстан*
[2]*ЖШС "Компра", 010000, Нұр-Сұлтан, Қазақстан*

Аңдатпа. *Берілген мақала ешбір тестілеусіз, адамның фотосуреті арқылы қандай MBTI типіне жататындығын анықтайтын шешімдердің бірін ұсынады. Instagram API көмегімен бұрын MBTI тестінен өткен респонденттердің жеке фотосуреттері жиналып, сол арқылы зерттеу жүргізілді. Жұмыс нәтижесінде дәлдіктің 60 пайызына қол жеткізілді.*

Түйінді сөздер: *жеке тұлғаны болжау, кескіндерді классификациялау, CNN, Instagram API, MBTI*

ПРОГНОЗ ТИПА MBTI С ИСПОЛЬЗОВАНИЕМ ИЗОБРАЖЕНИЙ ИЗ INSTAGRAM

МУСАЕВА Д.[1], КУАНДЫКОВА А.[2], ТАЛАСБЕК А.[1], ОРЫНБЕКОВА К.[1], КАРИБОЗ Д.[1]

[1]*Университет имени Сулеймана Демиреля, 040900, Каскелен, Казахстан*
[2]*ТОО "Компра", 010000, Нур-Султан, Казахстан*

Аннотация. *В данной статье описывается прогнозирование типа личности на основе изображений. С помощью Instagram API получили фотографии респондентов, ранее сдавших тест MBTI. В результате была достигнута 60-процентная точность. Эта работа представляет собой одно из решений, позволяющих определить, к какому типу личности принадлежит человек, не проходя никаких тестов, а просто загружая изображение.*

Ключевые слова: *предсказание личности, классификация изображений, CNN, Instagram API, MBTI*

Introduction

The Myers–Briggs Type Indicator (MBTI) - a questionnaire, which can determine psychological preferences of a person towards the world, other people and in decision making[1].

Scientists Krug and Johns did the MBTI experiment interrogating, approximately, 17000 people. They brought out six psycho points: Conscientiousness, Control, Tough Poise, Intelligence, Neuroticism and Extraversion. The results were independently confirmed by Boyle on the Cattell, Comrie and Eysenck scale [2].

E-I measurement defines extraversion and introversion. S-N measurement - Tough Poise (Sensitivity). T-F and J-P dimensions are Conscientiousness, Control, Intelligence, Openness, according to the Norman Big Five. MBTI focuses on cognitive characteristics, in contrast to other personality tools [3].

The MBTI is an opportunity to make the right decision while choosing a future profession, - say the Tiegers. Due to MBTI people can find the most appropriate job in which they can show maximum performance in building their career. They declare that personality type affects the quality of performing job responsibilities of chosen profession [4].

Implementation

One of the important stages in Machine Learning projects is the data collection part. Due to the lack of data, datasets from Kaggle and Google form that assembled in previous work were combined [5]. After scraping Images from Instagram accounts of responders more than 500 human faces were taken. To make the data collection process more accurate and faster, only the last 12 posts from each user were collected and only photos that are closer to the photo from the passport were chosen. The algorithm itself detects only the person's face and saves only this part as it is shown in Figure 1 below.
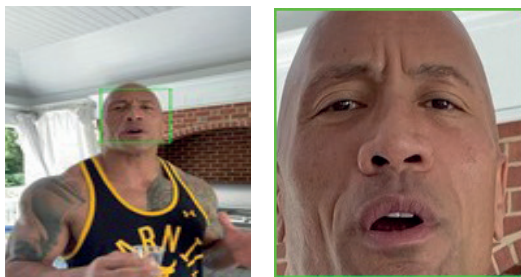

Figure 1. Image preprocessing [6]

Table 1. Custom dataset with personality type

Image	Type
cristiano.jpg	INTP
arianagrande.jpg	INFP
therock.jpg	ENFP
selenagomez.jpg	INFJ
kyliejenner.jpg	INFP
...	...
joannagaines.jpg	INFP
iza.jpg	INTP
serenawilliams.jpg	INFJ
ileana_official.jpg	INFP
irfanhakim75.jpg	INTP

Each image in our dataset shown in Table 1 was converted to Grayscale to avoid complexities, the colored image is in 3 dimensional array whereas gray is one only. To normalize the pixel values of an image, it was divided by 255, since this is the largest value that an array can take, to get a result from 0 to 1. For each pixel an array of three digits describe the color scheme: RGB (Red, Green, Blue).

Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is the most used neural network model for the task of classifying images [7]. The CNN spends less time and space, due to the fact that it takes not every pixel individually, but a certain square, then rotates the pixel that is most suitable for the criteria. Alternatively, fully connected weight network from each pixel, CNN has sufficient weights to look at a tiny piece of the image as shown in Figure 2.
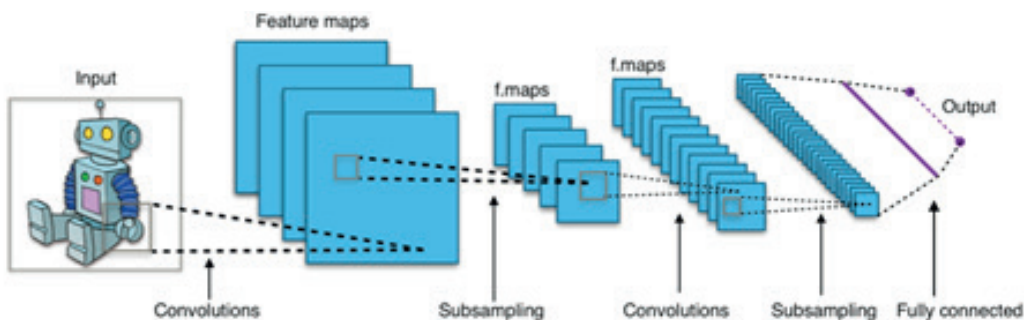

Figure 2. Convolutional Neural Network [8]

The shape of Input Layer 64x64 , three Convolutional layers with increasing filter size and left same values for kernel size as 5, with 1 stride length, padding 'same' and activation function 'ReLU' which is shown in Figure 3.
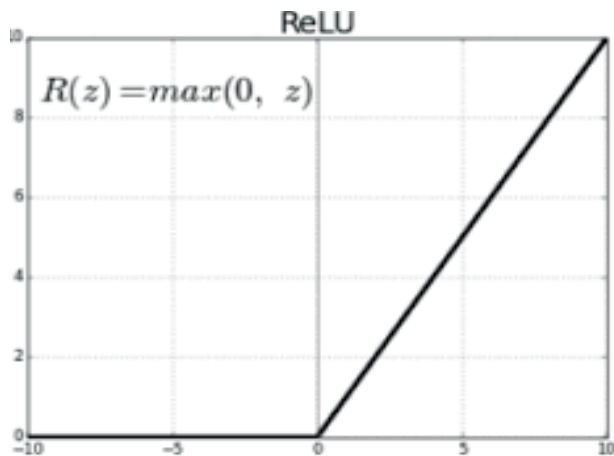

Figure 3. ReLU

Filter is a matrix for determining features from an image, it is a combination of connections. Filter size was increased because of features extraction from noisy data and each time combinations are getting complex, becomes abstract and hence we can work with larger size. Kernel size is the size of these convolutional filters, 5x5 square-shaped. Padding determines the size for output volumes based on input and to keep the same output volume size as the input is 'same padding', so it will add zeros around to the input volume. Immediately after the convolutional layer , the MaxPool2D pooling layer was used. In our case, it reduces the spatial dimension of output volume.

After convolutional layers we use one dropout layer to keep the network generalized and avoid overfitting data (caused by a complex model, performs well on training data but poor for unseen test data). To make classification we need fully connected layers without any complex structure, just a large piece of ready output data. To convert neurons of the convolutional layer it needs to be flattened, using the Dense layer to transfer 3-dimensional array to one dimensional. Softmax activation function that shown in Figure 4 is intended for categorical targets.
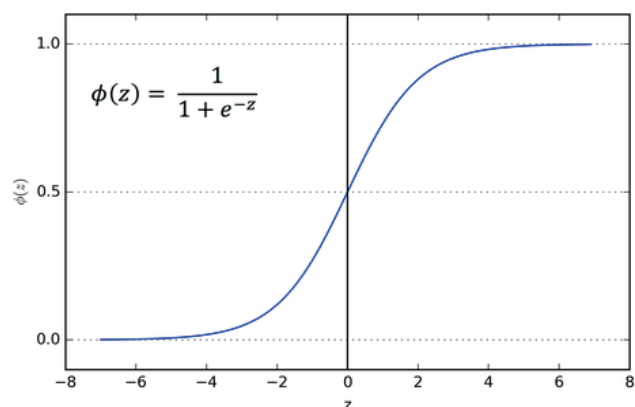

Figure 4. Softmax activation function

Optimizer determines how the learning proceeds and uses Adam optimizer with a learning rate of 0.001.

Results and conclusion

After the model is completed training, accuracy of 55.07% and 2.61 loss value on average were achieved. Figure 5 illustrates custom images in JupyterLab with the predicted MBTI type and accuracy of them.
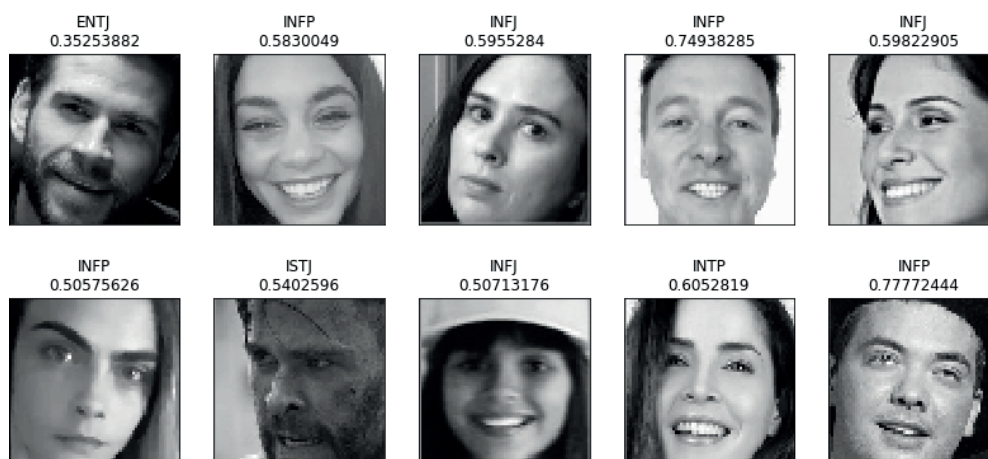

Figure 5. Results of datasets on jupyter

This article describes whether facial features can tell which personality type a person belongs to. As a result, 60 percent probability was achieved. CNN gives high performance in images as it has filters and "condition detectors". As with all work related to machine learning, the limitation of this work is lack of data. Another thing that should be noticed is that CNN gives as some results but doesn't give us analysis of features and how they affect on prediction itself.

REFERENCES

1. Myers, Isabel Briggs with Peter B. Myers (1995) [1980]. Gifts Differing: Understanding Personality Type. Mountain View, CA: Davies-Black Publishing. ISBN 978-0-89106-074-1
2. GJ. Boyle, (1989). Re-examination of the major personality-type factors in the Cattell, Comrey, and Eysenck scales: Were the factor solutions by Noller et al. optimal? Personality and Individual Differences, 10, 1289-1299.
3. P.T. Costa, & R.R. McCrae, (1992). NEO Personality Inventory. Psychological Assessment, 4, 5- 13.
4. Paul D. Tieger and Barbara Barron-Tieger. Do What You Are: Discover the Perfect Career for You Through the Secrets of Personality Type
5. Assem Talasbek, Azamat Serek. Personality Classification Experiment by Applying k-Means Clustering. International Journal of Emerging Technologies in Learning (iJET) 15(16):162.
6. Photo of actor. Available at: https://baigenews.kz/news/dueyn_dzhonson_zanyal_trete_mesto_v_mire_po_chislu_podpischikov_v_instagram/ Date of access [17.01.2020]
7. Nigam V. Understanding Neural Networks. From neuron to RNN, CNN, and Deep Learning (2018). Available at: https://towardsdatascience.com/understanding-neural-networks-from-neuron-to-rnn-cnn-and-deep-learning-cd88e90e0a90 Date of [17.02.2020]
8. CNN. Available at: https://towardsai.net/p/machine-learning/beginner-guides-to-convolutional-neural-network-from-scratch-kuzushiji-mnist-75f42c175b21 Date of access [17.01.2020]

Information about authors:

1. Mussayeva D. – B.S, Student of Master program, Suleyman Demirel University, Abylai Khan street 1/1, Kaskelen
2. Kuandykova A. – B.S, Data Analyst, JSC "Kompra", Mangilik Yel avenue, C4 /6, Nur-Sultan
3. Talasbek A. - MSc, Senior Lecturer, Suleyman Demirel University, Abylai Khan street 1/1, Kaskelen
   Email: assem.talasbek@sdu.edu.kz
4. Orynbekova K. – MSc, Senior Lecturer, Suleyman Demirel University, Abylai Khan street 1/1, Kaskelen
5. Kariboz D. – MSc, Senior Lecturer, Suleyman Demirel University, Abylai Khan street 1/1, Kaskelen