Recognition of Turkish Command to Play Chess Game Using CNN
Satranç Oyunu için CNN Kullanılarak Türkçe Komutları Tanıma

Yakup Kutlu[1,*], Gizem Karaca[2]
[1]Department of Computer Engineering, Iskenderun Technical University, Hatay, Turkey
[2]Department of Computer Engineering, Adana Alparslan Turkes University, Adana, Turkey
ORCIDs: 0000-0002-9853-2878, 0000-0001-7138-4097
E-mails: yakup.kutlu@iste.edu.tr, gkaraca@atu.edu.tr
*Corresponding author.

***Abstract*—A platform has been created that allows playing chess with Turkish voice commands. The aim of this study is to enable individuals with limited movement abilities as a result of congenital reasons or a certain disease or accident to play chess and perform a social activity without the help of another person, and to be rehabilitated at the same time. It consists of three parts: Chess module, Human-computer interaction module and Artificial Intelligence module. 29 words have been determined to provide movement in the game on the platform. Voice recordings from 151 people, 86 men and 65 women, were used. Feature selection was made on 43790 voice recordings by using mel frequency cepstral coefficients (MFCC) and gammatone cepstral coefficients (GTCC) methods. The results obtained were classified using the traditional CNN model. The data obtained after using MFCC and GTCC methods were used as inputs in the CNN model. In addition, the data obtained by the two methods were combined and trained in the model. Depending on the methods used in the created model, 83% to 85.9% results were obtained. It was determined that the results obtained using the MFCC method were more successful.**

***Keywords*—Chess; MFCC; GTCC; deep learning; human-computer interaction**

***Özetçe*—Türkçe ses komutları ile satranç oynanmasını sağlayan bir platform oluşturulmuştur. Bu çalışmanın amacı doğuştan, belirli bir hastalık ya da kaza sonucu hareket yetenekleri kıstlanmış bireylerin sosyal bir etkinlik olarak satranç oynamalarını ve başka bir kişinin yardımı olmadan sosyal bir aktivite gerçekleştirebilmelerini ve bir yandan da rehabilite olmalarını sağlamaktır. Satranç modülü, İnsan bilgisayar etkileşim modülü ve Yapay Zeka modülü olmak üzere üç bölümden oluşmaktadır. Platform üzerinde oyun içinde hareketin sağlanması için 29 sözcük belirlenmiştir. 86 erkek, 65 kadın olmak üzere 151 kişiden alınan ses kayıtları kullanılmıştır. 43790 ses kaydı üzerinde mel frekanslı cepstral katsayıları (MFCC) ve gammatone cepstral katsayıları (GTCC) yöntemleri kullanılarak öznitelik seçilimi yapılmıştır. Elde edilen sonuçlar geleneksel CNN modeli ile sınıflandırma işlemi yapılmıştır. CNN modelinde girdi olarak MFCC ve GTCC yöntemleri kullanıldıktan sonra elde edilen veriler kullanılmıştır. Ayrıca iki yöntem ile elde edilen veriler birleştirilerek model içinde eğitime alınmıştır. Oluşturulan model de kullanılan yöntemlere bağlı olarak %83 ile %85,9 sonuç elde edilmiştir. MFCC yöntemi kullanılarak elde edilen sonuçların daha başarılı olduğu belirlenmiştir.**

***Anahtar Kelimeler*—Satranç; MFCC; GTCC; derin öğrenme; insan-bilgisayar etkileşimi**

I. Introduction

It was desired to create a platform with the study where people with limited mobility skills can play chess with voice commands. In the study on chess other studies on chess development were examined [1]–[3].

Based on the interface, it was determined that 29 words should be recorded. Voice recordings were taken from 151 people, 86 men and 65 women. Data were collected by taking 10 records from the words determined from each person. Then, feature selection was made on these sounds with MFCC and GTCC methods. Later the classification process is performed with the developed CNN model [4]–[6].

By using MFCC and GTCC methods, accurate analysis and synthesis of audio signals are provided [7]. The automatic speech recognition interface created for the Macedonian language consists of a database containing 2.5 hours of data from 30 native languages and 188 words [8]–[10]. It is aimed to classify Tunisian and Moroccan dialects and determine the best method using the feedback back propagation neural network (FFBPNN) and SVM method [11]. In the study, in which LPC and GTCC methods were used for feature selection and classification with K-NN and HMM methods, a speech processing and recognition system was created for Myanmar language [12]. Since the results of Vietnamese voice commands with Google speech recognition (GSR) are quite low, a system with a higher success rate was created by using SVM and CNN models in the study [13]. In the system created by using 3 different CNN architectures, it is ensured that the Bangla language recognizes short speech commands [14]. Using the MFCC method, Vanilla Single-Layer Softmax Model, Deep Neural Network and Convolutional Neural Network

Models, a Speech Command Recognition system was created with Google TensorFlow and AIY team's Speech Command Dataset [15]. The system, which enables the operation of electronic devices with voice commands, performs real-time speech recognition in Kannada, one of the Indian languages. Quantum convolutional neural network (QCNN) is designed as a decentralized feature extraction method in speech recognition [16]. An in-depth study analysis of the operating performance of the CNN model has been carried out [17].

II. Materials & Methods

A. Data Acquisition

Voice recordings were taken from 151 people, 86 men and 65 women. Before the sound recordings were taken, 29 words were determined to be used depending on the interface. 10 records were taken from each of these determined words while recording from individuals. Afterwards, these sound recordings were pre-processed to get rid of unnecessary sound and noise.

B. Feature Extraction

The sound recordings, which have undergone pre-processing [18], [19], are first feature selected with the MFCC and GTCC methods. Here, the maximum features obtained for each method were determined as 4615 and 4603, respectively. These attributes are used as inputs in the CNN model.

C. Mel Frequency Cepstral Coefficients (MFCCs)

The MFCC method is one of the most frequently used methods for feature selection in audio signals. In this method, the audio signal is first divided into 20-30 second intervals by framing. Then, the windowing process is applied to eliminate the time difference between the frames. The Fourier transform is performed after the windowing process, and frequency-dependent processing is provided. Then, 12-14 cepstral features are obtained from an audio signal by performing Mel frequency filter and inverse Fourier transform operations. In our study, 13 cepstral features were obtained from each audio signal.

D. Gammatone Cepstral Coefficients (GTCCs)

GTCC method is a method based on MFCC method. In this method, the operations applied in the MFCC method are applied sequentially. However, unlike MFCC method, gammatone frequency filter is applied instead of mel frequency filter. The number of cepstal features obtained by this method is the same as the MFCC method.

E. Convolutional Neural Networks (CNNs)

The traditional CNN model was created. Data obtained by using MFCC and GTCC methods were used as inputs. However, since the dimensions of the sound recordings are different from each other after preprocessing, it has been determined that the dimensions obtained as a result of framing are different from each other. Padding was done so that the data could be used in the model. While doing this process, it has been tried two methods: first, making all sound recordings equal size, then applying MFCC and GTCC methods, or sound recordings going through MFCC and GTCC processes first and then reaching equal size. While these methods were applied for the MFCC process, two methods could not be applied for the GTCC process. It has been noticed that the GTCC data received without sizing is assigned as null. For this reason, two different data were created with the MFCC method and a data set with GTCC. The maximum dimensions obtained with MFCC and GTCC methods for a sound class were determined as 4615 and 4603, respectively. In addition, the data sets formed by combining the two data sets were trained in the model as input. While creating the model structure, a frame with dimensions of 71 x 65 x 1, 78 x 59 x 1, 71 x 130 x 1 was created, respectively. Then, the maximum and average pooling method was applied to the data, and the classification process is performed after the flatten layer. The results obtained with two different pooling methods were compared.

III. Experimental Study

In this study, the CNN model was created on Google Colab with the python software language. While creating the model, python_speech_features for feature selection and Conv2D, MaxPooling2D, AveragePooling2D libraries for CNN model were used. While creating the model, 43790 voice recordings and 5 data sets obtained by MFCC and GTCC methods were used as inputs. The size of the data set before feature selection by MFCC and GTCC methods was determined as 43790 x 57000. Raw data could not be used as input because the dimensions were too large. On top of that, the size of the data set was reduced by applying feature selection methods. While the data set for MFCC is 43790x4615, the data set for GTCC is determined as 43790x4603. The size of the data set formed by combining the obtained data sets was determined as 43790x9230. The reason for the creation of 5 data sets is that two different results are obtained by performing the padding process before or after the MFCC method is applied to the audio recordings. These results were trained on the model as two separate data sets during the merging process. While creating the model, two different results were obtained by using the MaxPooling2D and AveragePooling2D parameters. It was determined that the results obtained with the two parameters were close to each other.

IV. Results & Discussion

In the results obtained using the CNN model, 43790 voice recordings were used for the 29 words determined. The data obtained by using MFCC and GTCC methods were trained in the model and results were obtained (Table I). The results obtained are between 82.9% and 85.83%. It has been determined that the data sets in which the MFCC method is applied first and then sized are more successful. A result of 90.12% was obtained by classifying the voice recordings of 50 people (23 men, 27 women) for 29 words by the cross-validation method after the GTCC method [20]. When the CNN method and the Cross-validation method were compared, it was determined that the cross-validation method gave more successful results.

Features	AveragePooling2D	MaxPooling2D
MFCC first	85.8%	85.6%
MFCC last	83.2%	83.2%
GTCC only	83.2%	82.9%
MFCC fist + GTCC	85.8%	85.8%
MFCC last + GTCC	83.2%	83.2%

Table I: Classification accuracies using the CNN module with two pooling methods.

AUTHORS' CONTRIBUTIONS

This study is the part of G. Karaca and Y. Kutlu is the supervisor.

REFERENCES

[1] Janko V, Guid M. A program for progressive chess. Theoretical Computer Science 2016; 644: 76-91.

[2] Nabiyev VV. Providing harmonization Aamong different notations in chess readings. In 2011 IEEE 19th Signal Processing and Communications Applications Conference (SIU), April 20-22, 2011, Antalya, Turkey, pp. 29-33.

[3] Newell A, Shaw JC, Simon HA. Chess-playing programs and the problem of complexity. Book chapter in Computer Games I. Springer, New York, USA, 1988, pp. 89-115.

[4] Yildirim O, Ucar A, Baloglu UB. Recognition of real-world texture images under challenging conditions with deep learning. Journal of Intelligent Systems with Applications 2018; 1(2): 122-126.

[5] Narin A, Pamuk Z. Effect of different batch size parameters on predicting of COVID19 cases. Journal of Intelligent Systems with Applications 2020; 3(2): 69-72.

[6] Dervisoglu S, Sarigul M, Karacan L. Interpolation-based smart video stabilization. Journal of Intelligent Systems with Applications 2021; 4(2): 153-156.

[7] Fathima R, Raseena PE. Gammatone cepstral coefficient for speaker Identification. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering 2013; 2(1): 540-545.

[8] Gerazov B, Ivanovski ZA. A speaker independent small vocabulary automatic speech recognition system in Macedonian. In Proceedings of the Second International Conference (TAKTONS), November 13-16, 2013, Novi Sad, Serbia.

[9] Telceken M, Kutlu Y. Detecting abnormalities in heart sounds. Journal of Intelligent Systems with Applications 2021; 4(2): 137-143.

[10] Balli O, Kutlu Y. Regional signal recognition of body sounds. Journal of Intelligent Systems with Applications 2021; 4(2): 157-160.

[11] Hassine M, Boussaid L, Messaoud H. Maghrebian dialect recognition based on support vector machines and neural network classifiers. International Journal of Speech Technology 2016; 19(4): 687-695.

[12] Khaing I, Lin KZ. Automatic speech segmentation for Myanmar Language. International Journal of Scientific Engineering and Technology Research 2014; 3(24): 4726-4729.

[13] Nguyen QH, Cao TD. A novel method for recognizing Vietnamese voice commands on smartphones with support vector machine and convolutional neural networks. Wireless Communications and Mobile Computing 2020; 2020: 2312908.

[14] Sumon SA, Chowdhury J, Debnath S, Mohammed N, Momen S. (2018, September). Bangla short speech commands recognition using convolutional neural networks. In 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), September 21-22, 2018, Sylhet, Bangladesh, pp. 1-6.

[15] Li X, Zhou Z. Speech command recognition with convolutional neural network. CS229 Course Report, Stanford University, USA, 2017.

[16] Pavan GS, Kumar N, Krishna KN, Manikandan J. (2020, May). Design of a real-time speech recognition system using CNN for consumer electronics. In 2020 Zooming Innovation in Consumer Technologies Conference (ZINC), May 26-27, 2020, Novi Sad, Serbia, pp. 5-10.

[17] Huang JT, Li J, Gong Y. An analysis of convolutional neural networks for speech recognition. In 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), April 19-24, 2015, South Brisbane, QLD, Australia, pp. 4989-4993.

[18] Sayilgan E, Yuce YK, Isler Y. Frequency recognition from temporal and frequency depth of the brain-computer interface based on steady-state visual evoked potentials. Journal of Intelligent Systems with Applications 2021; 4(1): 68-73.

[19] Degirmenci M, Sayilgan E, Isler Y. Evaluation of Wigner-Ville distribution features to estimate steady-state visual evoked potentials' stimulation frequency. Journal of Intelligent Systems with Applications 2021; 4(2): 133-136.

[20] Karaca G, Kutlu Y. Turkish voice commands based chess game using gammatone cepstral coefficients. Journal of Artificial Intelligence with Applications 2020; 1(1): 1-4.