Polina Shakhova, Yevheniia Kolomiiets, ArtemVolokyta

METHOD BASED ON CONVOLUTIONAL NEURAL NETWORK FOR MUSICAL CHORD RECOGNITION

The article deals with the issue of Automatic Chord Recognition (ACR) using Convolutional Neural Networks (CNN). Recognition is based on the Mel-Frequency Cepstral Coefficients (MFCC) of the input audio signal. The proposed system allows to recognize 25 basic chords. Python language and TensorFlow library were used for the development.

Keywords: Automatic Chord Recognition, Convolutional Neural Networks, MFCC, TensorFlow.

Relevance of the topic. The problem of the automatic chord recognition has been known since the last century. Automatic chord recognition systems are widely used in many areas, for example for music generation, for classification of the music into various categories (genre, mood), for song identification, etc.

Target setting. Nowadays there are many conferences, competitions and other events in the field of Music Information Retrieval (MIR), including the annual ISMIR conference to exchange ideas and innovations related to this field, MIREX competition of algorithms for musical chords recognition, etc. However, there is no unambiguous and completely optimal solution of this problem at the moment and the existing solutions need improvements.

Actual scientific researches and issues analysis. A lot of research in this area has focused on deep neural networks: convolutional (CNN) [4, 5], recurrent (RNN) [2], long short-term memory (LSTM) [7].

Uninvestigated parts of general matters defining. The following issues were noticed in reviewed works:

- training datasets usually consist of songs of one artist or one genre, that can affect the model's quality of recognition on the real data [1, 2];
- model input is usually audio features, that require to store a large amount of data [2, 4, 7];
- some of the models is able recognize only a small number of chords [3].

The research objective. The purpose of this paper is to investigate the application of the convolutional neural networks to recognize musical chords from music audio recordings. As a solution, the article focuses on creating a model that takes into account the previous observations. Proposed model is able to classify 25 types of chords: 12 minor and 12 major chords and 1 "non-chord" type. A set of Mel-Frequency Cepstral Coefficients or MFCC extracted from each audio signal is used as model input features. The model is implemented using Python and TensorFlow library.

The statement of basic materials. In general, the task of recognizing song chords can be defined as: for a given sound signal $x(t)$, where $t \in [t_{start}, t_{end}]$ and a set of possible chord classes Y, for each moment of time t it is necessary to determine the chord that sounds at that moment.

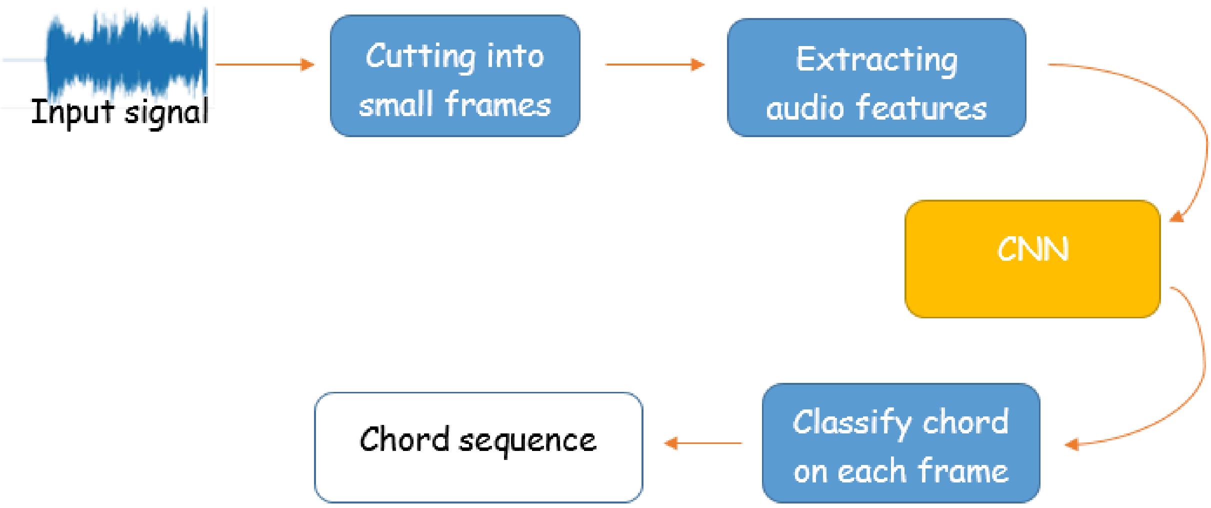

Figure 1. General system structure

Model input features. MFCC were decided to use as model input features, because unlike the other sound characteristics they represent a small amount of data, providing a lot of information at the same time.

Creating a model. A convolutional neural network architecture was chosen to perform this task. Convolutional Neural Networks (CNNs) are widely used in image classification, but they also have shown very good results in audio processing (speech recognition, music identification).

Table 1

Configuration of the proposed CNN

Convolution	N × 40 × 32 × 3 layers
Max pooling	3 × 3
Convolution	9 × 20 ×64 × 2 layers
Max pooling	3 × 3
Convolution	5 × 10 × 128 × 2 layers
Max pooling	2 × 2

Flatten	1920
Fully-connected	128
Softmax	25

The output of each convolutional layer is activated with the ReLU function. Batch normalisation is performed after two first max pooling layers. Dropout with probability 0.3 is applied after the second batch normalization layer, last max pooling layer and fully-connected layer.

Experiments. In addition to the Isophonics collection of 180 transcribed *The Beatles'* songs, which is traditionally used for chord recognition tasks, the training set in this work is extended with also popular Isophonics *Queen* dataset and a collection of popular rock and pop songs presented at one of the stages of MIREX competition. The total dataset consists of 379 songs with their transcriptions stored in the format: *Start_time – End_time – Chord*

	Start	End	Chord
0	0.000000	1.053119	N
1	1.053119	3.593854	B:min
2	3.593854	6.090000	G
3	6.090000	8.655804	E
4	8.655804	11.140340	A
5	11.140340	13.659705	A
6	13.659705	16.109410	C#:min
7	16.109410	18.686825	F#:min
8	18.686825	19.371814	D

Figure 2. Example of chord transcription file

Chords in each song transcription file were simplified to 25 basic chord classes. Both audio and chord transcription files were aligned by time and cut into small (about 0,2 seconds) frames. After extracting MFCCs from each audio frame, results were saved into JSON-file.

A large range of input data can affect neurons and lead to incorrect adjustment of coefficients, so the training data was additionally aligned before feeding into the neural network. This means reducing the data to the interval [-1; 1].

Evaluation the results. The main metrics for evaluating the results of training NN are accuracy function and loss function. Loss function shows the average cost of training for the epoch, and accuracy function - the amount of correctly classified data.

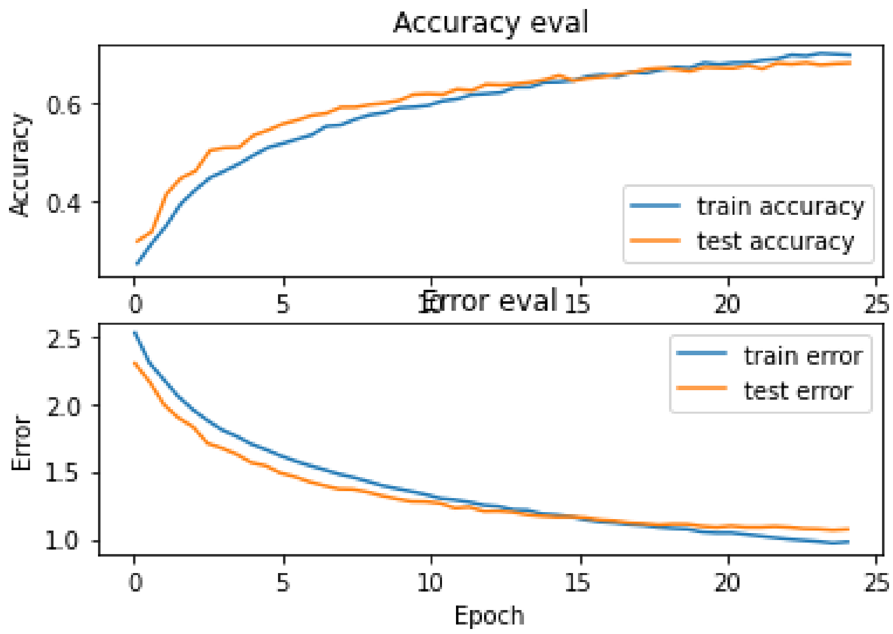

Figure 3. Loss and accuracy function results

Throughout the learning process, the amount of loss is constantly decreasing, while the accuracy is increasing. Both graphs for training and test data are similar in shape and change synchronously.

Conclusions. This paper demonstrates the method based on Convolutional Neural Networks for Automatic Chord Recognition tasks. Proposed model is able to recognize all basic minor and major triads and “non-chords” from a set of Mel-Frequency Cepstral Coefficients extracted from each input audio signal. It can be seen that the use of such combination produces good results. The model was trained on extended dataset of the songs, so it is expected to perform qualitative recognition on the real data.

The possible direction for future work is to increase the number of chords that the model is able to recognize. This task is primarily complicated by the uneven use of different chords and could require a more complex neural network architecture.

References

1. H.-T. Cheng, Y.-H. Yang, Y.-C. Lin, I.-B. Liao, and H. H. Chen. (2008). *Automatic chord recognition for music classification and retrieval.* In 2008 IEEE International Conference on Multimedia and Expo. (pp. 1505–1508).
2. N. Boulanger-Lewandowski, Y. Bengio, and P. Vincent. (2013). *Audio chord recognition with recurrent neural networks.* In ISMIR. (pp. 335–340).
3. J. Osmalskyj, J-J. Embrechts, S. Piérard, M. van Droogenbroeck. (2012). *Neural networks for musical chords recognition.* In Journées d’Informatique Musicale, hal ID: hal-03041758.

4. F. Korzeniowski, G. Widmer. (2016, Sept. 13–16). *A fully convolutional deep auditory model for musical chord recognition*. In IEEE International Workshop on Machine Learning for Signal Processing.
5. E. J. Humphrey, J. P. Bello. (2012). *Rethinking automatic chord recognition with convolutional neural networks*. In 11th International Conference on Machine Learning and Applications IEEE.
6. M. McVicar, R. Santos-Rodriguez, Y. Ni, T. D. Bie. (2014). *Automatic Chord Estimation from Audio: A Review of the State of the Art*. In IEEE ACM Transactions on Audio, Speech, and Language Processing. (pp. 556 – 575).
7. S. Nakayama and S. Arai. (2018). *DNN-LSTM-CRF model for automatic audio chord recognition*. In Proceedings of the International Conference on Pattern Recognition and Artificial Intelligence (pp. 82–88).

AUTHORS

Shakhova Polina – bachelor, Department of Computer Engineering, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”.

E-mail: schachowa.paulina@gmail.com

Kolomiiets Yevheniia – bachelor, Department of Computer Engineering, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”.

E-mail: jennichka10155@gmail.com

Volokyta Artem – associate professor, Department of Computer Engineering, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”.

E-mail: artem.volokita@kpi.ua