Group A9 report

INSTRUMENT CLASSIFICATION IN MUSIC

Members:
Mathias Plans,
Ali Jafarov,
Alexandra Elsakova

https://github.com/mathiasplans/instrument-classifier

Business understanding

Today, many areas of life are becoming more and more computerized due to the intensive development of the IT-sphere, which makes our existence not only easier, but also more productive. In this regard, we decided to make a Tool that can be used for various purposes combining with testing ourselves in the new field (working with sound), which was not which was not covered in the course "Introduction to Data Science". This project is carried out not only to complete the work within the training course, but also solves a number of problems. First of all, it can be used in kindergartens and schools for teaching primary school children. What is more it can be applied in work with people who have lost their hearing or were born deaf. Last but not the least is to make searching from the audio database easier. The project is considered successful if, upon completion, we would obtain at least 90% of the accuracy of our model.

As the main resources we have 3 people with experience in different fields (1 bachelor student from Information Technology curricular, 2 master students from Bioengineering with experience in general biology and medical engineer), 3 computers with decent graphics cards, 1 database (size 11 GB) and also Python with libraries which are required.

We plan to finish the work by December 13th and present the project at a poster session, which is on the 17th of December, 2020. We have no legal or security obligations. The work takes place in real time, the required result is creating an AI model which has ~ 90% accuracy at least. We also require labeled data.

The work identified the following risks:

1. Wrongly labeled data, which can be corrected by manually working through the information.
2. Terrible audio quality, which is solved by excluding bad samples or using sufficient amounts of data.
3. Unbalanced data, which can be solved by balancing information.

This work is an open source project, so the final tool has no monetary benefits. The only costs that are required in our work are human resources, electricity, internet, essentials for life during the "creative creation process".

Our Data-mining goals are a Database of features (processed from initial dataset), AI model (s) that classify instruments, and a poster needed to present the work and summarize it. Data-mining success criteria is Model performance is 90% accuracy at least, and supervisors' satisfaction. Equally important is our own gratification after teamwork and project presentation.

Terminology (based on [1]):

- *Temporal Features* are the features related to the shape of the envelope of a node (includes temporal moment, temporal centroid, temporal width, temporal asymmetry, temporal kurtosis).
- *Spectral Features* are characteristics, associated with the fact that timber is the colour of sound or tone, are called spectral characteristics (includes spectral moment, Spectral Slope, spectral Decrease, Spectral Roll-off, Spectral Flux, Audio Spectral Flatness, Spectral crest Factor).
- *Cepstral Features* are commonly used to characterize speech and music signals.

- *Perceptual Feature* is a central construct of perception.
- *Zero Crossing Rate* is a measure of the number of times in a given time interval/frame that the amplitude of the speech signals passes through a value of zero.
- *Auto-Correlation* is the correlation of a signal with a delayed copy of itself as a function of delay.
- *Constant Q transform* is a data series to the frequency domain.
- *Auto-regressive* is a representation of a type of random process.
- *Mel-Frequency Cepstral Coefficients* are coefficients that collectively make up an MFC.
- *Total Loudness and Specific Loudness* are the distribution of loudness.
- *Sharpness* is a hearing sensation related to frequency and independent of loudness.
- *Signal-to-mask ratio* is the distance between the level of the masker and the masking threshold.
- *MultiLayer Perceptron* is a class of feedforward artificial neural networks.
- *Polyphony* is a property of musical instruments that means that they can play multiple independent melody lines simultaneously.
- *Monophony* is a synthesizer that produces only one note at a time.

[1] Gulhane S. R., Badhe S. S., Shirbahadurkar S.D. Cepstral (MFCC) Feature and Spectral (Timbral) Features Analysis for Musical Instrument Sounds. IEEE Global Conference on Wireless Computing and Networking (GCWCN). 2018. P. 109-113.

Data understanding

For this project, the Irmas database [2] is used, which includes 11 GB of files (musical fragments + text description). The instruments considered are: cello, clarinet, flute, acoustic guitar, electric guitar, organ, piano, saxophone, trumpet, violin, and human singing voice. Occurring files have both polyphonic and monophonic sound. For work, fragments of the required length will be selected (at the present time it is quite difficult to talk about this, since the music in them can be slow (then you can miss a number of musical instruments or meet a pause [3]). Data can be MP3 or WAV files. We are using the sound file, and the data that is in the title of the files (which include what instruments are playing).

The dataset consists of multiple genres. We are only interested in the classical genre (labelled 'cla'). The predominant instruments in the genre 'cla' are:

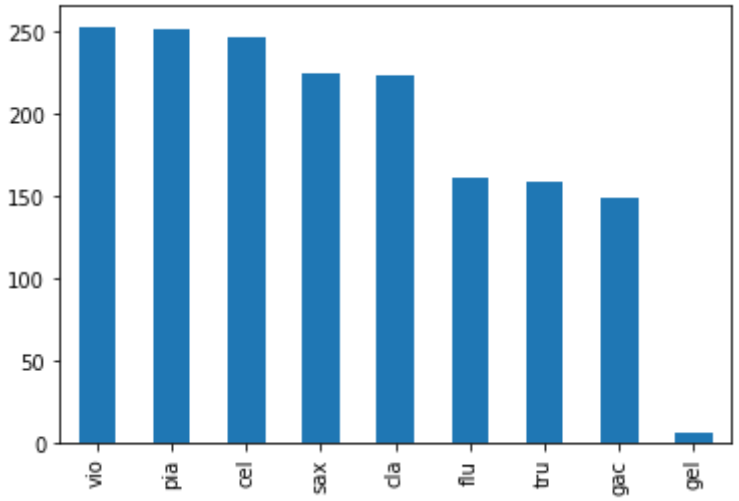

Table 1. Instrument distribution in the classical genre

The instruments that are least included are electric guitar ('gel') and acoustic guitar ('gac'). Flute ('flu') and trumpet ('tru') are also underrepresented compared to others. If we want the dataset to be balanced, then each instrument can have around 150 samples (if we exclude electric guitar). This means that the training data size is 8 * 150 = 1200 samples. Unbalanced dataset would have 1675 samples. Each sample is 3 seconds long, which means that the instruments that are playing in one segment do not change much during the piece.

The IRMAS dataset does not label all the instruments that are playing in a segment. Only predominant instruments, genre, and presence of drums are labeled. If we want to identify all the instruments that are playing at all times, we need to add labels to the dataset ourselves. This means relabeling 1200 to 1675 samples. After the relabeling, this dataset will suit us perfectly.

[2] Bosch, J. J., Janer, J., Fuhrmann, F., & Herrera, P. "A Comparison of Sound Segregation Techniques for Predominant Instrument Recognition in Musical Audio Signals", in Proc. ISMIR P. 559-564. 2012.

[3] Beigi H. Audio Source Classification using Speaker Recognition Techniques. P. 1-6. 2011.

Plan

Work on this project began a month ago, so the planning task is almost complete. We have also partially completed task 3. The entire plan is shown below.

Task ID	Description	Mathias	Ali	Alexandra
1	Plan the project, write documentation	8h	8h	11h
2	Set up a database, scripts	2h	-	-
3	Find or implement functions for processing the data (getting the features) (find libraries)	8h	12h	10h
4	Create a new database with processed data	3h	-	-
5	Use K-NN, Gaussian Mixture Classifier (or any other simpler model) for classification	2h	6h	6h
6	Use some kind of neural network (NN) for classification	10h	4h	-
7	Create a poster	30m	2h	4h
Total hours spent		33.5h	32h	31h

In our work we plan to use: Neural networks, AI, K-NN and other algorithms (if K-NN is not the best solution). The platform for our work is Jupyter Notebooks with several libraries (scipy, matplotlib, numpy, Keras, etc.). We use SLACK, WhatsApp, Discord as the communication platform with additional materials shared in Google Drive (URL: https://drive.google.com/drive/folders/16cVGi5YIPqaJaeNksLh7NklvlDtMB2N7?usp=sharing) and GitHub.