Predominant Musical Instrument Classification based on Spectral Features

Harish(55) Ankit(41) Bhushan(59)
Vineet(46) Karthikeya(32)

Indian Statistical Institute

CDS Course Project Presentation
November 25, 2019

Musical Information Retrieval

Music information retrieval (MIR) is the interdisciplinary science of retrieving information from music.

Music information retrieval (MIR) is the interdisciplinary science of retrieving information from music.

Key areas in MIR

- Recommender Systems
- Track separation
- Genre Detection

Musical Information Retrieval

Music information retrieval (MIR) is the interdisciplinary science of retrieving information from music.

Key areas in MIR

- Recommender Systems
- Track separation
- Genre Detection
- Music Transcription
- Music Categorization
- Music Generation

Music information retrieval (MIR) is the interdisciplinary science of retrieving information from music.

Key areas in MIR

- Recommender Systems
- Track separation
- Genre Detection
- Music Transcription
- Music Categorization
- Music Generation

Musical Instrument Recognition

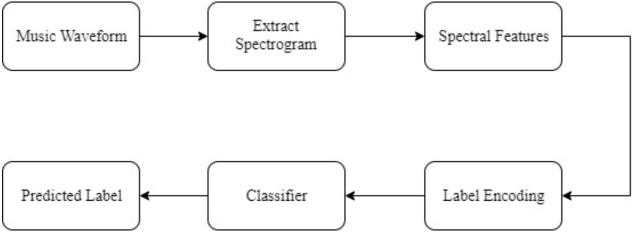


Figure: Workflow for Musical Instrument Identification

Dataset

IRMAS

annotated polyphonic dataset[1] with predominant musical instrument

upf. Universitat Pompeu Fabra Barcelona

MTG Music Technology Group

25 1994 2019

[1] Janer *et al.* ISMIR 2012

Dataset

annotated polyphonic dataset[1] with predominant musical instrument

Instruments

cello, clarinet, flute, acoustic guitar, electric guitar, organ, piano, saxophone, trumpet, violin, and human voice

[1] Janer *et al.* ISMIR 2012

Dataset

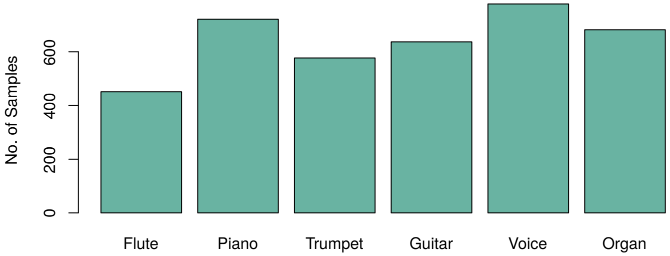
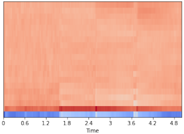


Figure: Number of audio samples per instrument class

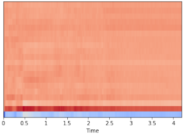
Total 3 hr 12 min of Audio Samples

Timbre

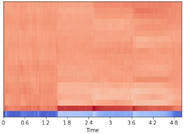
Timbre is the ‘colour’ of a sound. Timbre can distinguish between different types of string instruments, wind instruments, and percussion instruments.



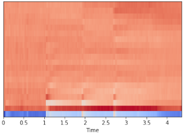
(a) EDM



(b) Guitar (Accoustic)



(c) Key Board



(d) Organ

Figure: Same note (audio) played on various instruments

MFCC Calculation Flowchart

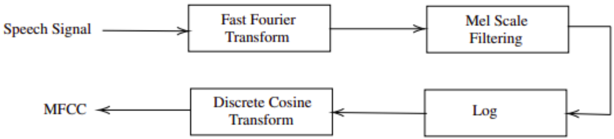


Figure: MFCC Calculation Workflow

MFCC Calculation Flowchart

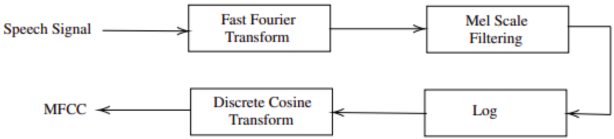


Figure: MFCC Calculation Workflow

	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	-0.714190	0.462601	-0.633688	-2.809177	-1.932569	-1.929835	-1.273824	-3.261680	-2.245337	-1.041374	1.420384	0.030057	-2.085074	flu
1	-0.597550	0.812213	-0.877901	-2.620451	-1.904552	-2.578117	-2.140579	-2.119974	-0.606116	-0.714572	-0.824773	-1.153546	-0.675475	flu
2	-0.393161	0.457444	-0.857359	-3.008680	-2.101997	-1.788754	-1.219133	-3.484734	-2.775243	-0.966307	1.904729	0.021335	-2.818950	flu
3	-0.701514	-3.845421	-2.317507	-1.570193	0.052099	2.576265	1.182365	-0.268957	-0.239296	-2.318107	2.599968	4.855833	-0.369089	flu
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
255	-1.588576	1.861871	1.551132	-0.497802	-0.289868	-1.639274	1.401908	1.066984	0.201532	-1.121848	-1.044130	0.045466	0.587872	flu

256 rows × 14 columns

```
np.mean(mfcc-feature-vector,axis=1)
```

MFCC Calculation Flowchart

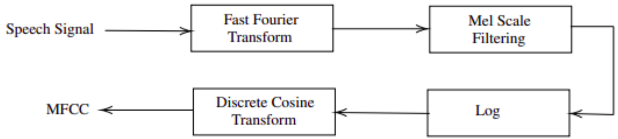


Figure: MFCC Calculation Workflow

	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	-0.714190	0.462601	-0.633688	-2.809177	-1.932569	-1.929835	-1.273824	-3.261680	-2.245337	-1.041374	1.420384	0.030057	-2.085074	flu
1	-0.597550	0.812213	-0.877901	-2.620451	-1.904552	-2.578117	-2.140579	-2.119974	-0.606116	-0.714572	-0.824773	-1.153546	-0.675475	flu
2	-0.393161	0.457444	-0.857359	-3.008680	-2.101997	-1.788754	-1.219133	-3.484734	-2.775243	-0.966307	1.904729	0.021335	-2.818950	flu
3	-0.701514	-3.845421	-2.317507	-1.570193	0.052099	2.576265	1.182365	-0.268957	-0.239296	-2.318107	2.599968	4.855833	-0.369089	flu
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
255	-1.588576	1.861871	1.551132	-0.497802	-0.289868	-1.639274	1.401908	1.066984	0.201532	-1.121848	-1.044130	0.045466	0.587872	flu

256 rows × 14 columns

```
np.mean(mfcc-feature-vector,axis=1)
```

0	1	2	3	4	5	6	7	8	9	10	11	12	13
-1.407441	-0.144181	-1.200199	-1.520351	-0.093229	-0.213987	-0.346466	-0.927328	-0.365662	-0.313958	0.527216	1.032585	0.208176	flu

Other Spectral Features

- Zero Crossing Frequency — simple measure of the frequency content of a signal
- Root mean Square — rms summarises the energy distribution of each frame
- Spectral Centroid — It is a measure of average frequency weighted by the sum of spectral amplitude within one frame
- Spectral Bandwidth — frequency range of a signal weighted by its spectrum
- Spectral Rolloff — measure of rolloff frequency

Methodology

- We experimented with two libraries – Essentia & Librosa for feature extraction
- Each audio sample of 3 sec produced 257 rows. We took mean and produced a single vector per audio file
- Labelled each vector using `labelencoder`
- trained the classifier in `scikit learn`

Supervised Classification

- Logistic Regression — (Baseline Model)
- Decision Tree
- LGBM
- XG Boost
- Random Forest
- Support Vector Machine

- Precision is the ratio $\frac{tp}{(tp+fp)}$ where tp is the number of true positives and fp the number of false positives. The precision is intuitively the ability of the classifier not to label as positive a sample that is negative.

Model Evaluation

- Precision is the ratio $\frac{tp}{(tp+fp)}$ where tp is the number of true positives and fp the number of false positives. The precision is intuitively the ability of the classifier not to label as positive a sample that is negative.
- Recall is the ratio $\frac{tp}{(tp+fn)}$ where tp is the number of true positives and fn the number of false negatives. The recall is intuitively the ability of the classifier to find all the positive samples.

Model Evaluation

- Precision is the ratio $\frac{tp}{(tp+fp)}$ where *tp* is the number of true positives and *fp* the number of false positives. The precision is intuitively the ability of the classifier not to label as positive a sample that is negative.
- Recall is the ratio $\frac{tp}{(tp+fn)}$ where *tp* is the number of true positives and *fn* the number of false negatives. The recall is intuitively the ability of the classifier to find all the positive samples.
- F1 score can be interpreted as a weighted average of the precision and recall.

$$F1 = \frac{2 \times (\text{precision} * \text{recall})}{(\text{precision} + \text{recall})}$$

Model Evaluation

- Precision is the ratio $\frac{tp}{(tp+fp)}$ where *tp* is the number of true positives and *fp* the number of false positives. The precision is intuitively the ability of the classifier not to label as positive a sample that is negative.
- Recall is the ratio $\frac{tp}{(tp+fn)}$ where *tp* is the number of true positives and *fn* the number of false negatives. The recall is intuitively the ability of the classifier to find all the positive samples.
- F1 score can be interpreted as a weighted average of the precision and recall.

$$F1 = \frac{2 \times (\text{precision} * \text{recall})}{(\text{precision} + \text{recall})}$$

- Confusion Matrix is a technique to evaluate performance of a supervised classification. Calculating a confusion matrix gives a better idea of what our classification model is getting right and what types of errors it is making.

Accuracy Statistic

	Logistic Regession			Decision Tree			LGBM		
Instrument	P	R	F1	P	R	F1	P	R	F1
Flute	0.58	0.39	0.47	0.43	0.44	0.43	0.66	0.59	0.62
Piano	0.55	0.59	0.57	0.53	0.54	0.53	0.69	0.73	0.71
Trumpet	0.44	0.53	0.48	0.50	0.46	0.48	0.59	0.67	0.63
Guitar	0.63	0.57	0.60	0.60	0.57	0.58	0.73	0.68	0.71
Voice	0.58	0.48	0.52	0.52	0.50	0.51	0.72	0.54	0.62
Organ	0.51	0.61	0.56	0.50	0.55	0.52	0.63	0.74	0.68

	XG Boost			RF			SVM		
Instrument	P	R	F1	P	R	F1	P	R	F1
Flute	0.66	0.59	0.62	0.72	0.48	0.58	0.63	0.63	0.63
Piano	0.72	0.71	0.71	0.72	0.75	0.74	0.79	0.84	0.81
Trumpet	0.58	0.69	0.63	0.61	0.72	0.66	0.78	0.77	0.78
Guitar	0.71	0.72	0.71	0.73	0.72	0.72	0.77	0.76	0.77
Voice	0.75	0.53	0.62	0.74	0.54	0.62	0.78	0.67	0.72
Organ	0.65	0.74	0.69	0.63	0.80	0.70	0.79	0.85	0.82

Table: Precision, Recall & F1 Score for various Supervised Models

Model Evaluation – F1 Score

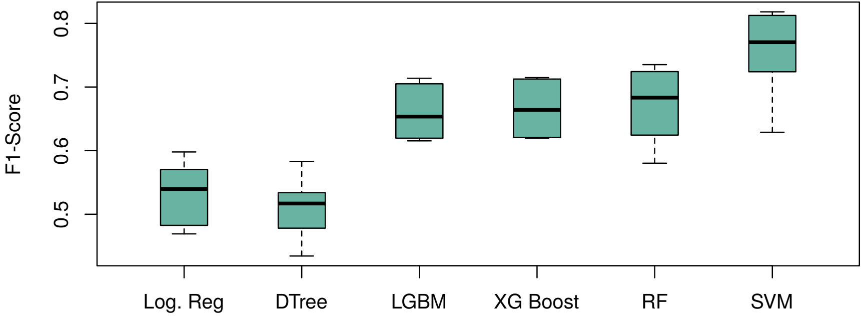


Figure: F1 Measure for Various Models

Instrument Performance

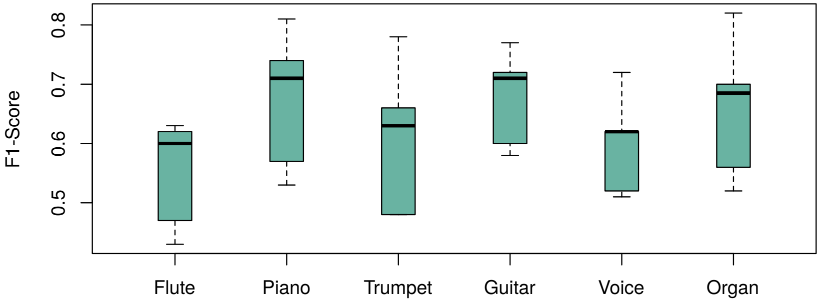


Figure: Instrument wise classification

Model Evaluation – Confusion Matrix

(a) Logistic Regression

(b) Decision Tree

(c) Light GBM

Figure: Confusion Matrix for various supervised Algorithms

(a) XG Boost **(b)** Random Forest **(c)** SVM

Figure: Confusion Matrix for various supervised Algorithms

Unsupervised Approach

K-means Clustering

Hierarchical Clustering

NN Arch

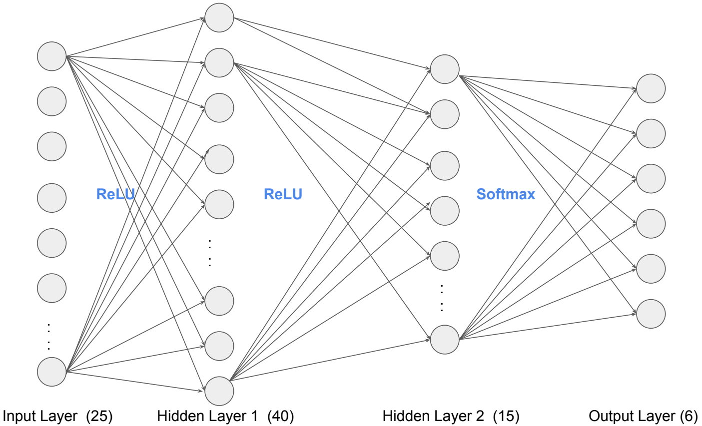


Figure: 3 Layer Neural Network

Loss Function: Cross Entropy
Minimizer Function: adam

Acknowledgement & References

- Bosch *et al.* A comparison of sound segregation techniques for predominant instrument recognition in musical audio signals. **ISMIR 2012**
- Deng *et al.* A study on feature analysis for musical instrument classification. **IEEE Transactions on Systems, Man, and Cybernetics 2008**
- Eronen *et al.* Musical instrument recognition using cepstral coefficients and temporal features. **ICASSP 2000**

All the code used is available in github.
`https://github.com/vntkumar8/musical-instrument-classification`

Thanks to intel Software Developer Zone

Thank You!
Questions?