Tabla Bol Recognition project

Blazej Kotowski and Ninad Puranik

Introduction

The tabla is a pair of drums known as ‘tabla’ and ‘bayan’ commonly used in accompaniment and solo performances in Indian Music. A variety of sounds can be produced using different kinds of strokes on the tabla-bayan. A mnemonic syllable or bol is associated with each of these strokes. We address the problem of recognizing the bol performed in an audio recording of a tabla solo performance with a harmonium accompaniment. We use a supervised learning approach to classify the ‘bols’ using the timbral features of audio and the manual annotations from the compmusic tabla-solo dataset.

Dataset

https://compmusic.upf.edu/tabla-solo-dataset
Dataset is well organized and documented. Onsets of tabla bols seem to be accurately marked and annotated. Mapping of similar sounding bols with different names is accurately provided. The dataset reflects a close to real life scenario due to the harmonium accompaniment.

The audio files are all 44100 fps 16-bit PCM Mono. This makes it difficult to do work with the separate tabla or tabla mixed with harmonium. Distribution of bols across categories is not uniform. Some bols are not sufficiently represented.

Previous work

S. Gupta et.al. [1] have developed the dataset comprising audio recordings, scores and time aligned syllabic transcriptions for 38 tabla solo compositions of different forms in tīntāl (a metrical cycle of 16 time units). The compositions are from the instructional video DVD Shades Of Tabla by Pandit Arvind Mulgaonkar, who is among the most renowned contemporary tabla maestros. They have approached the problem using a standard speech recognition model with HMM based pattern recognition. Gillet [2] and Chordia [3] have worked with a dataset with pure tabla sounds without harmonium accompaniment. GIllet has used power spectra of sounds while Chordia has used 31 manually selected timbral features (which include MFCCs)

The code developed by S. Gupta et. al is accessible at https://github.com/swapnilgt/percPatternDiscovery. Code could not be installed and executed as per the instructions provided in the github. The code may be complete and reproducible but was not easy for the authors to fully understand in the available time-frame for this work. This was one of the reasons to adopt a different methodology rather than extend the work of by S. Gupta et. al.

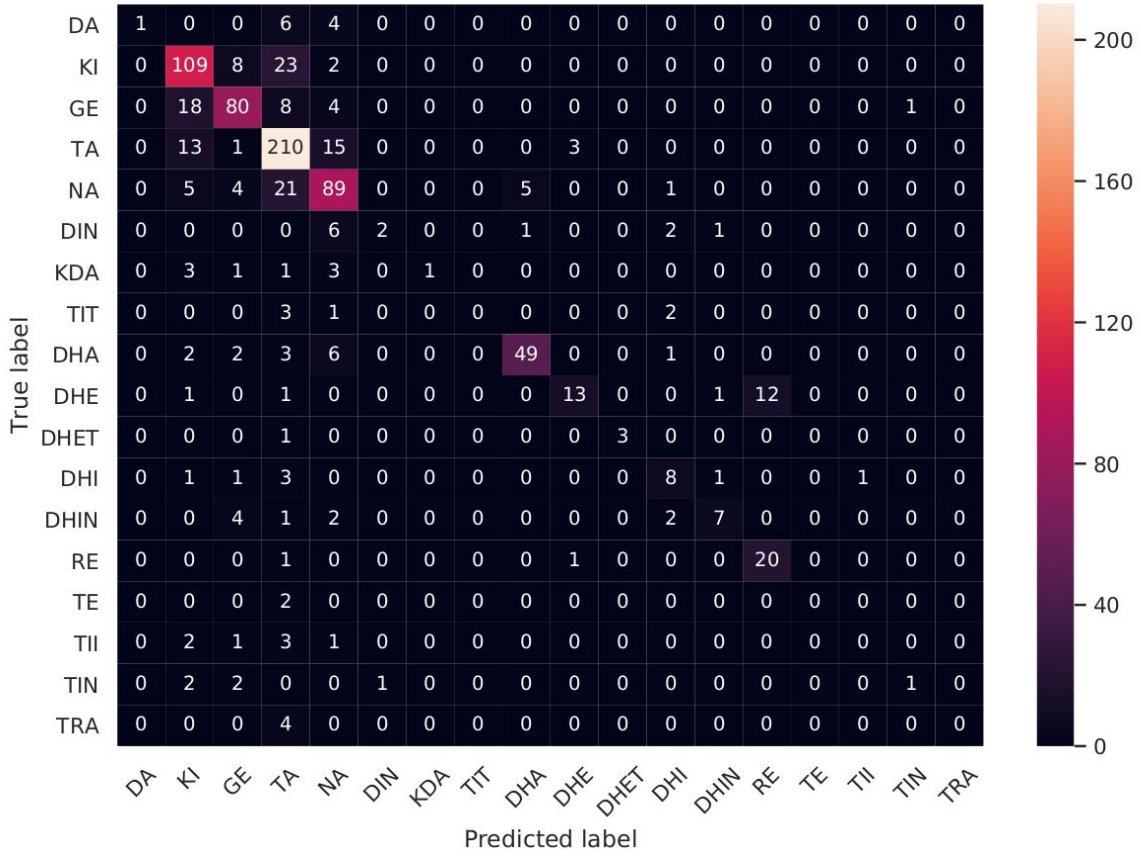
Methodology

We approach the problem differently than Gupta. We are extracting large number of features and then training different classifiers to get the best model for bol transcription. The feature extraction is done using essentia. In addition to the all low level features, we also extract the first and second derivatives of them to help us capture time varying spectral changes.

1) Collect 100 ms slices of audio starting from each annotated onset representing a bol.
2) For each audio slice extract variety of features. Use essentia extractor for this purpose. Extract HTK model based MFCC features, as described in [4], using our own extractor.
3) Create a new dataset with the ground-truth 'bol' annotation from the tabla-solo-dataset and the extracted features.
4) Train a classifier for this dataset for bol-recognition task.
5) Identify best features using feature selection models and obtain best possible classification, as measured by standard classification metrics like 'correctly classified instances' and ' precision, recall and F-measure'.

Results

We are presenting metrics for our best performing classifier and selection of features, which is Support Vector Machine with HTK-style MFCCs + derivatives.



	precision	recall	f1-score	support
DA	1.00	0.09	0.17	11.00
KI	0.70	0.77	0.73	142.00
GE	0.77	0.72	0.74	111.00
TA	0.72	0.87	0.79	242.00
NA	0.67	0.71	0.69	125.00
DIN	0.67	0.17	0.27	12.00
KDA	1.00	0.11	0.20	9.00
TIT	0.00	0.00	0.00	6.00
DHA	0.89	0.78	0.83	63.00
DHE	0.76	0.46	0.58	28.00
DHET	1.00	0.75	0.86	4.00
DHI	0.50	0.53	0.52	15.00
DHIN	0.70	0.44	0.54	16.00
RE	0.62	0.91	0.74	22.00
TE	0.00	0.00	0.00	2.00
TII	0.00	0.00	0.00	7.00
TIN	0.50	0.17	0.25	6.00
TRA	0.00	0.00	0.00	4.00
micro avg	0.72	0.72	0.72	825.00
macro avg	0.58	0.42	0.44	825.00
weighted avg	0.71	0.72	0.70	825.00

The rest of results can be accessed at this report.

Conclusions

- SVM is the best performing classifier for defined task.
- Speech-oriented HTK parameters for MFCC extraction perform surprisingly well for tabla classification problem.
- Adding temporal features (MFCC derivatives) may be very important for correct timbral classification.
- Spectral features are those that matter the most in case of tabla bol recognition

Future Work

- **Feature extraction**: Harmonium could be removed from the audio. This could possibly raise the quality of extracted features for tabla bols.
- **Classification**: Feature selection techniques could be further explored.
- **Transcription**: We could use onset detection coupled with this trained model to do an end-to end tabla transcription.

Extracted features

Classifiers can be run using the following pre-extracted features:

- All low-level features + HTK MFCCS + derivatives (only mfccs)
- All low-level features + HTK MFCCS + derivatives (all features)
- Extracted htk-MFCC features
- extracted features and HTK-MFCCs +derivatives starting 1 frame before onsets

References

[1] S. Gupta, A. Srinivasamurthy, M. Kumar, H. A. Murthy, X. Serra. Discovery of Syllabic Percussion Patterns in Tabla Solo Recordings. In Proc. of the 16th International Society for Music Information Retrieval Conference (ISMIR), 2015. http://compmusic.upf.edu/ismir-2015-tabla

[2] O. Gillet and G. Richard. Automatic labelling of tabla signals. In In Proc. of the 4th ISMIR Conf., 2003.

[3] P. Chordia. Segmentation and recognition of tabla strokes. In Proc. of the 6th International Conference on Music Information Retrieval (ISMIR), pages 107–114, London, UK, September 2005. http://ismir2005.ismir.net/proceedings/1137.pdf

[4] Young, S. J., Evermann, G., Gales, M. J. F., Hain, T., Kershaw, D., Liu, X., … Woodland, P. C. (2009). The HTK Book (for HTK Version 3.4). Construction, (July 2000), 384, https://doi.org/http://htk.eng.cam.ac.uk

[5] Github repository with the code: https://github.com/PNinad/tabla_bol_recognition_MIR_Project/