Polyphonic Music Transcription

A Deep Learning Approach

Aniruddha Zalani & Ayush Mittal

Department of Computer Science and Engineering

Course Porject
CS365 Artificial Intelligence
Guide - Amitabha Mukerjee

Abstract

In polyphonic music, many notes are played at once. Transcribing notes from the polyphonic music can help in plagirism detection, artist identification, Genre Classification, Composition Assitance and Music Tutoring Systems. Since, many notes are played at once, therefore, the techniques of multi class classification are not applicable here. In this project, we have learned 88 binary classifier which helps in transcribing notes of polyphonic music. Each classifier detects the presence of one note in the music at every time step. Unsupervised feature learning using RNN-RBM (Recursive Neural Networks and Restricted Boltzmann Machine). SVM classifiers are build using one-vs-all classification. HMM smoothing has been done to improve the results.

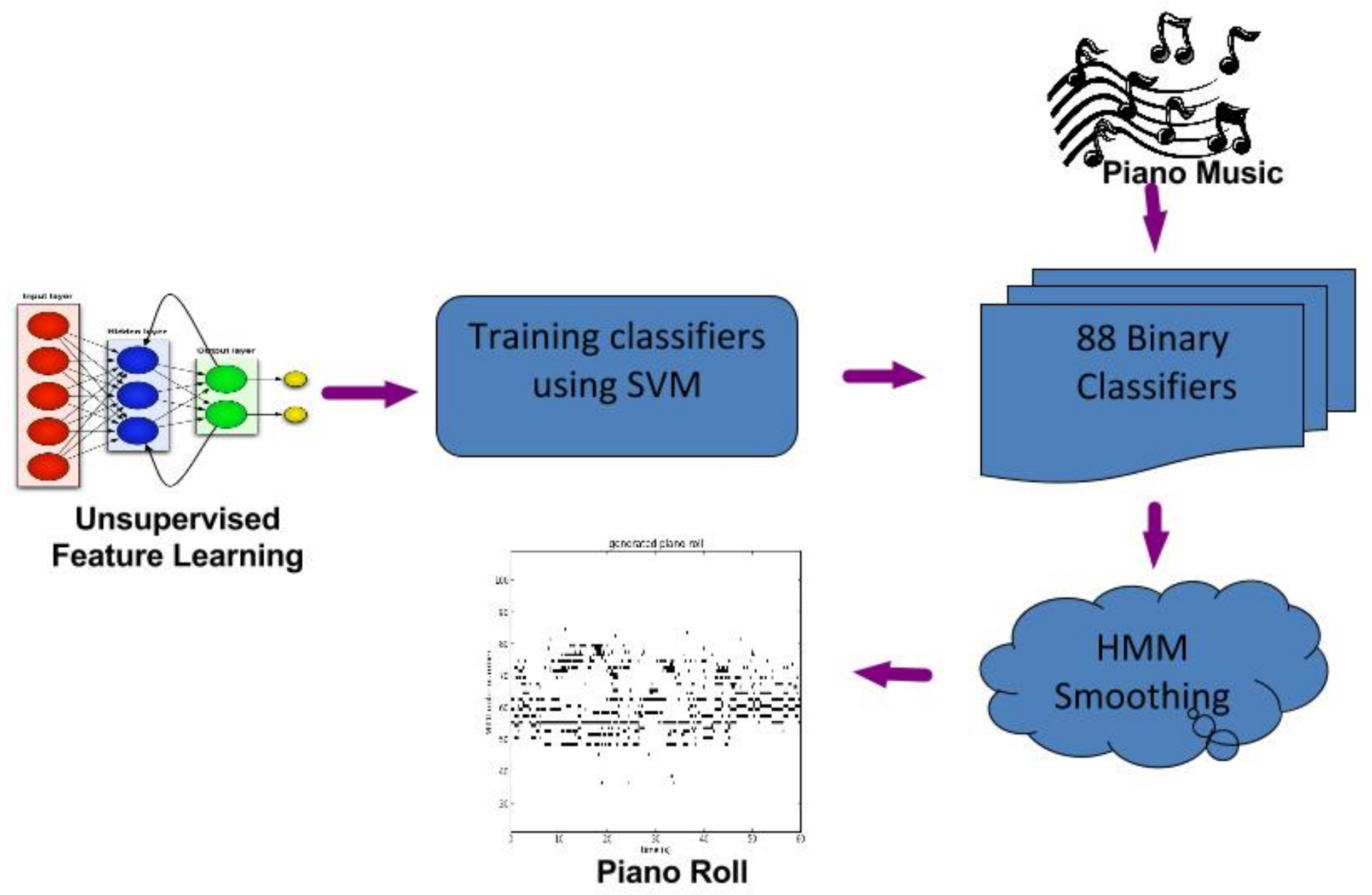
Introduction

1. In music, polyphony is a texture consisting of two or more simultaneous lines of independent melody.
2. More naturally ocuuring phenomena such as music, speech, are inheretly sequential.
3. Many notes are played at once, therefore, techniques of multi class classification are not applicable.
4. Some interesting work has been done using non-negative matrix factorization method[1] [2].
5. Most of the recent work involved use of deep learning methods for unsupervised feature learning.
6. Our approach is based on works of Nicholas et al.,[3] for feature learning.
7. For classification we have used Poliner and Ellis SVM based one-vs-all classification.

Main Objectives

1. Experiment with various deep learning methods such as Restricted Boltzmann Machine based Recurrent Neural Networks for unsupervised feature learing.
2. Build a classification model to extract the notes played in a polyphonic piano and tabla song.
3. Learn a classifier for each note.
4. Resythesize the song from the notes transcribed using these classifiers.

Methodology



Feature Learning

For feature learning we have experimented with two approaches: RNN-RBM based model.

RNN-RBM based model

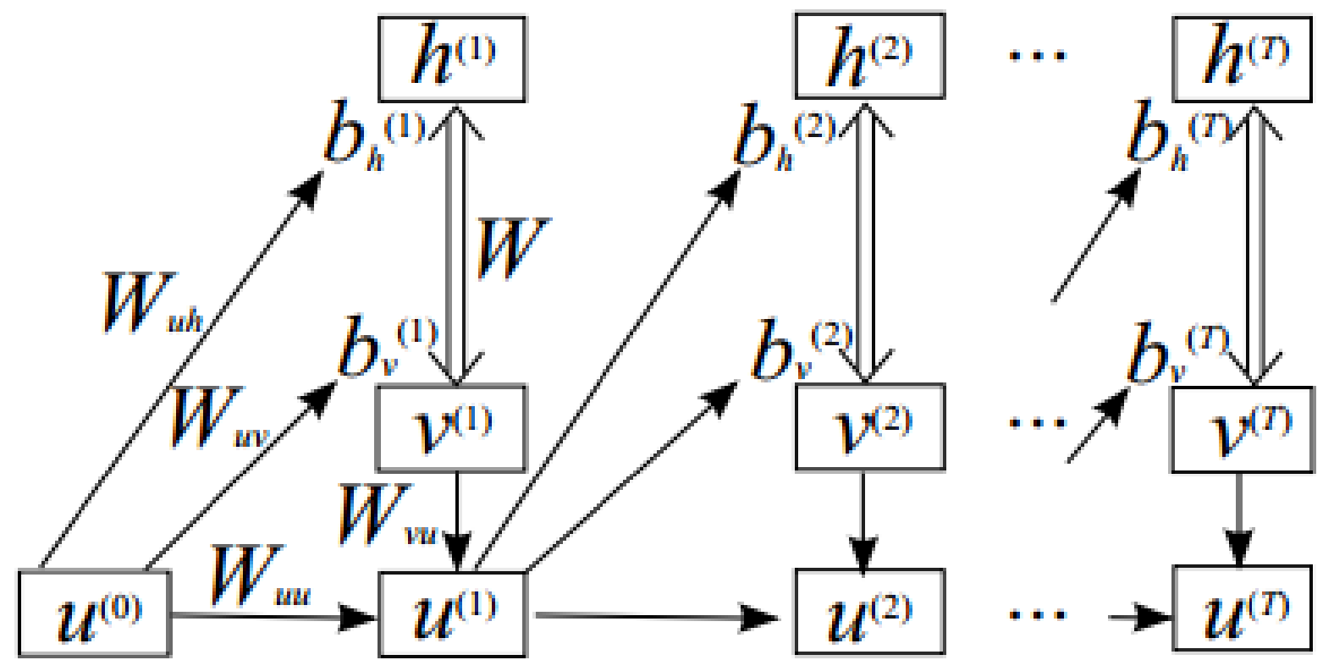


Figure 1: RNN-RBM unrolled over time

- It is an energy based model used for density estimation for temporal sequences.
- Multimodal Conditional distribution of v(t) given A(t) where $A^{(t)} = \{v_\tau | \tau < t\}$

Features

- Calculated STFT spectrogram.
- Concatenated them into a matrix.

Classification

1. In classification step along with features learned from unsupervised learning we have also used spectograms as features.
2. The classification is one-vs-all classification.
3. 88 binary SVM classifiers (linear kernel) are trained independently.

Smoothing

- Data from SVM is noisy and contains many spurious notes.
- Hidden Markov Model is used for smoothing the output of classification.
- Forward backward algorithm
- Advantage - Avoids estimation of probabilities to be zero, even for events never observed in the data.

Dataset

We have used a subset of MAPS dataset. Training data comprised of 6 piano files, with nearly thirty minutes of music. Test data comprised of 4 files with nearly 15 minutes of music. We have used a separate cross validation for each SVM classifier.

Results

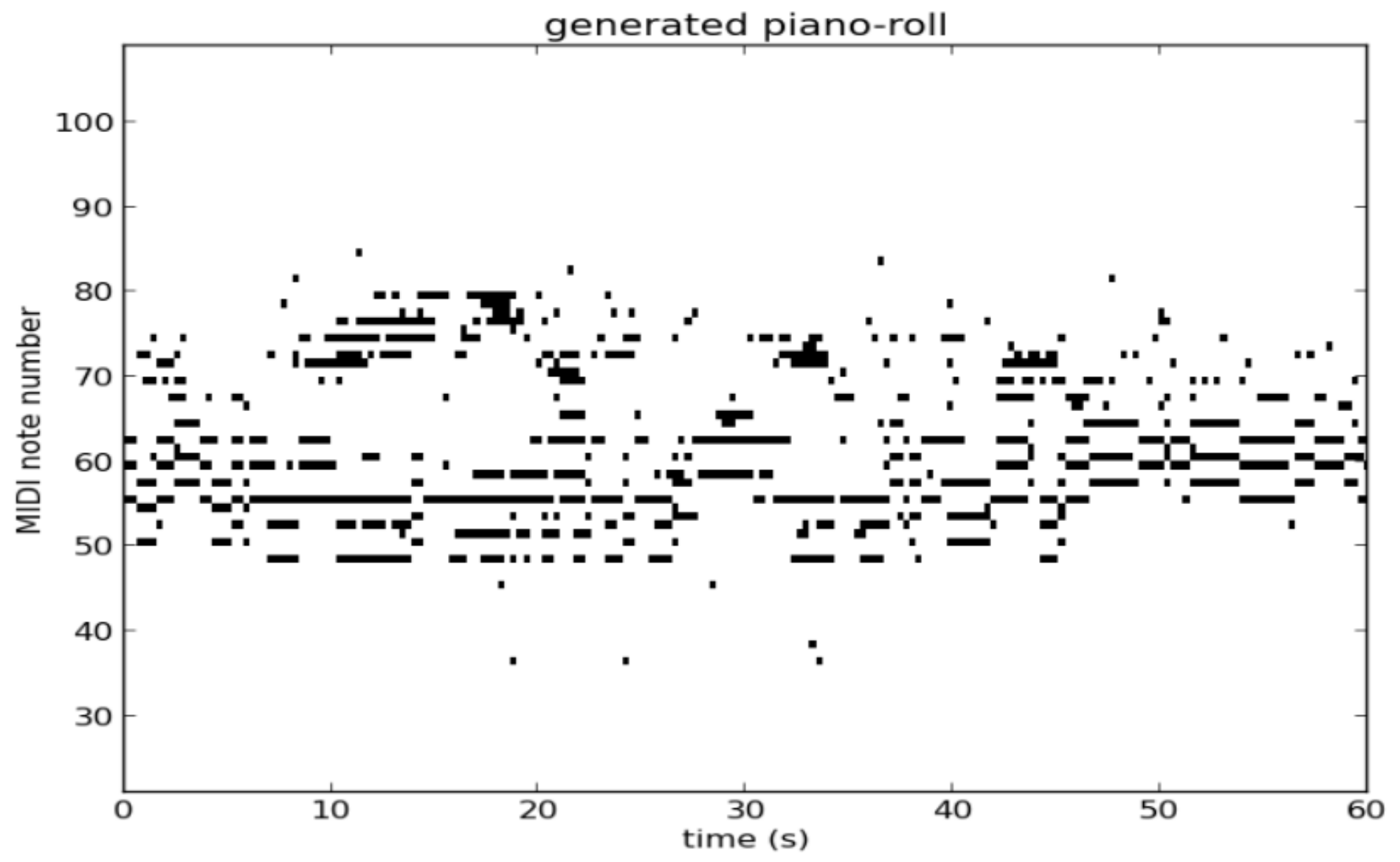


Figure 2: Features

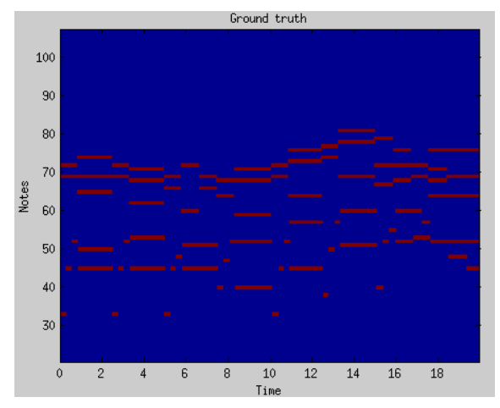


Figure 3: Ground

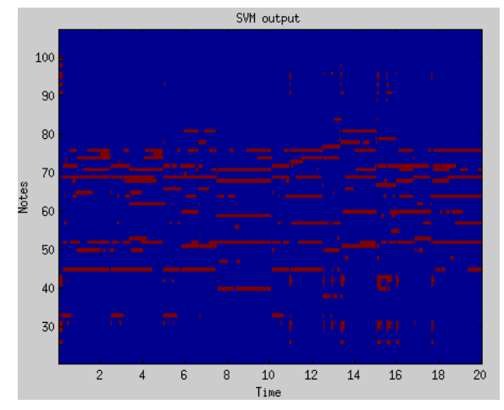


Figure 4: SVM

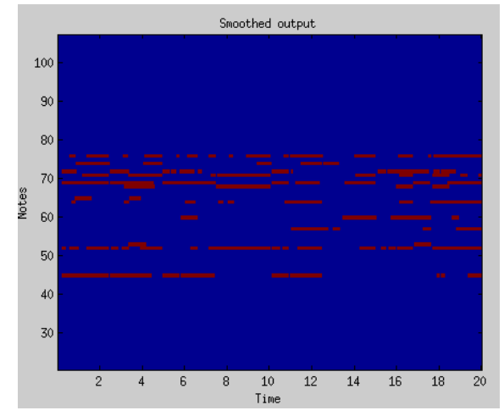


Figure 5: Smooth

Evaluation Metric

- Frame Level Accuracy - TP/(TP + FN + FP)
- Frame-level transcription error score (R_{total})
- E_{sub} number (at each frame) of ground truth notes for which some other note was reported,
- E_{miss} number of ground truth notes which cannot be accounted for.
- E_{fa} the number of reported notes which cannot be paired with a ground truth note.

Data	Accuracy	E_{tot}	E_{sub}	E_{miss}	E_{fa}
Smooth	0.6310	0.5426	0.1838	0.1570	0.2018
Raw	0.5207	0.8337	0.0951	0.0105	0.7281

Table 1: Results for piano

Algorithm	Accuracy
Polinear and Ellis	0.6770
RNN-RBM (Our Approach)	0.6310
Marolt [6]	0.396
Ryyananen and Klapuri [5]	0.4630

Table 2: Comparison from other techniques

Experiments with Tabla

We collected the Tabla dataset from different websites and we have learnt features from it using RNN-RBM but because of alignment problems in correspoding wav and midi files intermediate tabla roll was poor.

Conclusions

We have presented an unsupervised feature learning based approach with SVM classification and HMM smoothing for polyphonic music transcription. Our RNN-RBM based model achieves the accuracies close to state of art techniques. We have achieved 63.1 percent accuracy on piano dataset. Unsupervised feature learning improves results over simple Poliner-Ellis model. We have also experimented with tabla dataset. We also tried to learn features using convolutional deep belief network.

Future Work

The feature learning step can be improved through efficient implementations of Convolutional Deep Belief Networks(CDBN). The work can be further extended to various other music devices such as tabla and drums. Also the classification step can be made more efficient by using multi-note training instead of single note training.

References

[1] Arnaud , Arshia et al. Real-Time Detection of Overlapping Sound Events with Non-Negative Matrix Factorization

[2] Paris and Judith Non-Negative Matrix Factorization for Polyphonic Music Transcription, IEEE 2003

[3] N. Boulanger-Lewandowski, Y. Bengio and P.Vincent, Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription," ICML, 2012.

[4] J. Nam, J. Ngiam and H. Lee,Classication- Based Polyphonic Piano Transcription Approach Using Learned Feature Representations," ISMIR , pp. 175-180, 2011

[5] M. Ryynanen and A. Klapuri: Polyphonic music transcription using note event modeling, Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2005.

[6] M. Marolt: A connectionist approach to automatic transcription of polyphonic piano music, IEEE Transactions on Multimedia, vol.6, no.3, pp.439449, 2004.

Acknowledgements

We would like to thank Prof. Amitabha Mukerjee for guiding us throughout the project. We are very thankful to Nicholas Boulanger-Lewandowsk for providing tutorial for Recursive Neural Network. We are also thankful to Jerremy Nash for keeping codes of Poliner-Ellis Method as open source on github.