Music Stream Splitting

Vikul Gupta, Nidhi Manoj, and Sandhini Agarwal
CS 230, Stanford University 2018

MedleyDB & Problem Definition

- MedleyDB is a collection of 108 songs with melody annotations. 50 songs have isolated drums streams which are used to train our neural network model
- Splitting a mixed audio (wav) music file of instruments and vocal in order to isolate the vocal stream

Challenges

- Splitting up music into a variable number of instruments/ sources is an extremely difficult problem and not many research papers have had success
- Simply feeding in time steps into an LSTM causes vanishing gradient issues
- Need to compute a measure of error: we compare the spectrogram of the predicted drums and true drums audio files.

Approach: Baseline (Classifier)

- Built a logistic regression model that would classify songs as piano vs. not piano
- Achieved 80% accuracy

Approach: ICA

- Built an ICA model that would separate the song into two channels
- The first channel output corresponded to the instrument channel and the second one to the vocals channel.
- Compared difference between the channels of the input song and the output song

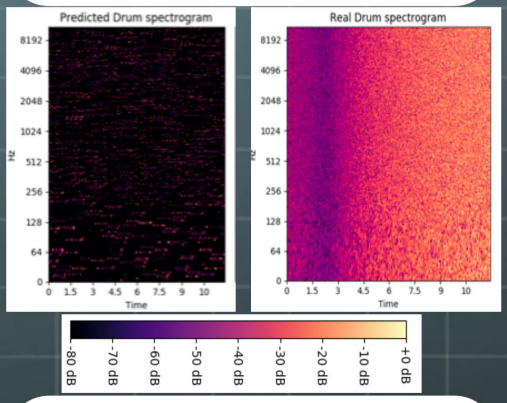
Evaluation Metric

- Summed the difference between drum spectrum (b) and our output (a)

$$\sum_j \sum_i \frac{|a_{ij} - b_{ij}|}{a_{ij}}$$

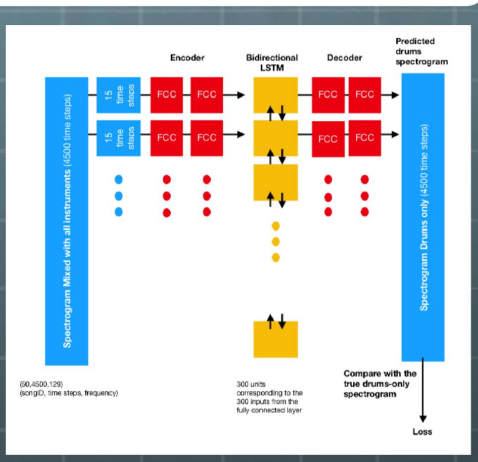
Approach: Fully Connected Layers & LSTM

- Found the spectrogram for each song file for 2580 timesteps and got input dimensions of **35*2580*513**
- Encoded by feeding these into two fully connected layers in **43 chunks**
- Passed layers output into a bi-directional LSTM of **60 units**
- Decoded the output of LSTM by passing it through two fully connected layers



Results and Analysis

- Our accuracy went from 0.0061 in the first epoch, to 0.0727 in the tenth epoch
- Our error rate reduced from ~2390 to ~1632 (after scaling both by 10^15)
- We think that one reason the model may be giving us a lower accuracy may be due high similarity between the drum channel and the mixed song
- We aim to test out this model with vocals to see if we get better results



References

- Logisticregression.http://scikit-learn.org/stable/modules/generated/sklearn.linear_mode 2017.
- R. Bittner, J. Salamon, M. Tierney, M. Mauch, C. Cannam, and J. P. Bello. Medleydb: A multitrack dataset for annotation-intensive mir research. 15th International Society for Music Information Retrieval Conference, Taipei, Taiwan, 2014.
- Michael Galarnyk. Logistic regression using python.
- https://towardsdatascience.com/logistic-regression-using-python-sklearn-numpy-mnist-
- 2017.
- marl. medleydb. https://github.com/marl/medleydb, 2017.
- . J.R. Simpson, G. Roma, and M. D. Plumbley. Deep karaoke: Extracting vocals from musical mixtures using a convolutional deep neural network. arxiv, 2015.
- [6] T. Virtanen. Unsupervised learning methods for source separation in monaural