Roberto Noel
Gus Xia
Independent Study: MIR
11/12/19

Progress Report 4: Neural Network, Labeled Data & Clustering Insights

These past two weeks I have been focusing on creating a "fitness" measure. To do this, I rated a dataset of 1000 randomly generated / web-scraped compositions on a scale from 1 to 100. While rating, I paid close attention to the song's melodic and rhythmic features. I have also found someone to re-rate these songs for additional input. This is highly subjective but at the time it is the best way to achieve labeled data. In the future, I plan to crowd-source more ratings.

I have decided to use a neural network based on both spectral and tree based features. This will allow me to train a model based on several features without running into the curse of dimensionality. So far I have trained a neural network on solely spectral features which has yet to yield satisfying results. I think we will see more promising results once the tree based features are added to the neural network and more data is labeled.

The current spectral and temporal features are:

- tempo: Tempo
- avg_cent: Average Spectral Centroid
- std_cent: Standard Deviation of Spectral Centroid
- avg_rolloff: Average Spectral Rolloff
- std_rolloff: Standard Deviation of Spectral Rolloff
- avg_zcross: Average Zero Crossing Rate
- std_zcross: Standard Deviation of Zero Crossing Rate
- avg_flt: Average Spectral Flatness
- std_flt: Standard Deviation of Spectral Flatness
- avg_bw: Average Spectral Bandwidth
- avg_ctr: Average Spectral Contrast
- std_ctr: Standard Deviation of Spectral Contrast

These features were all pulled from a Python library called Librosa. I am also working to extract structural features with a library called MSAF, though it might not fit within the scope of this semester.

Roberto Noel
Gus Xia
Independent Study: MIR
11/12/19

The current tree features I am working with are:

- depth: depth of tree
- operator_ct: count of each operator
- root: operator at root node
- operand_ct: count of non-“t” leaves
- t_ct: count of “t” leaves
- leaf_avg: average of non-“t” operands
- leaf_std: standard deviation of non-t operands

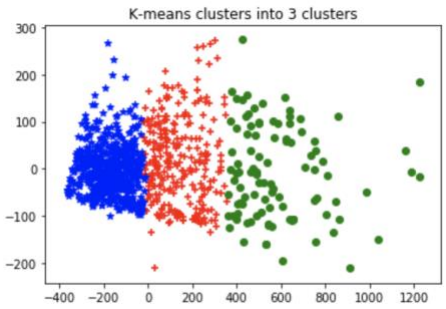
Due to the high subjectivity of the labeling, results from the neural network still remain poor, but if we can gain any sense of intuition as to which songs are better than others, we can start testing the genetic algorithm. The only remaining function that needs to be implemented for generation is crossover.

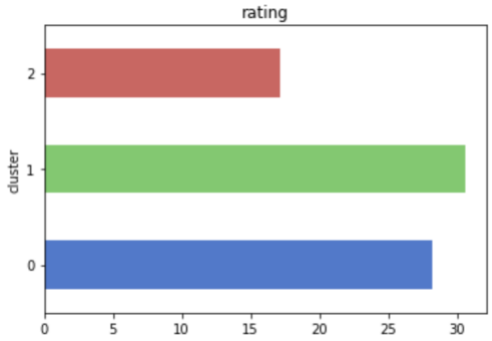
For crossover I will use insights from k-means clustering on both spectral and temporal features. To do this, I will only choose features with high correlations to the labels for clustering. This is necessary to avoid the curse of dimensionality. As we can see below, std_cent, std_rolloff, std_zcross and std_flt have the highest correlations with rating.

	tempo	avg_cent	std_cent	avg_rolloff	std_rolloff	avg_zcross	std_zcross	avg_flt	std_flt	avg_bw	avg_ctr	std_ctr
tempo	1	-0.154943	0.0319906	-0.12572	0.128588	-0.136103	-0.101615	-0.0733331	-0.0638587	-0.0380664	-0.0200012	-0.0293729
avg_cent	-0.154943	1	0.0992715	0.933313	-0.346741	0.876275	0.496006	0.42228	0.368693	0.687213	-0.0073753	-0.0777166
std_cent	0.0319906	0.0992715	1	0.0668535	0.759427	0.161578	0.722088	-0.0165592	0.0882293	0.0198781	0.0787752	0.20158
avg_rolloff	-0.12572	0.933313	0.0668535	1	-0.338707	0.730782	0.409429	0.332845	0.31217	0.860939	-0.00103031	-0.093876
std_rolloff	0.128588	-0.346741	0.759427	-0.338707	1	-0.205163	0.276805	-0.218385	-0.123041	-0.260301	0.114903	0.236835
avg_zcross	-0.136103	0.876275	0.161578	0.730782	-0.205163	1	0.56708	0.383697	0.284318	0.475173	-0.00685699	-0.0163799
std_zcross	-0.101615	0.496006	0.722088	0.409429	0.276805	0.56708	1	0.218644	0.285604	0.248421	0.00320195	0.043057
avg_flt	-0.0733331	0.42228	-0.0165592	0.332845	-0.218385	0.383697	0.218644	1	0.735266	0.118301	-0.317891	-0.296883
std_flt	-0.0638587	0.368693	0.0882293	0.31217	-0.123041	0.284318	0.285604	0.735266	1	0.139993	-0.204228	-0.152919
avg_bw	-0.0380664	0.687213	0.0198781	0.860939	-0.260301	0.475173	0.248421	0.118301	0.139993	1	-0.0148839	-0.13145
avg_ctr	-0.0200012	-0.0073753	0.0787752	-0.00103031	0.114903	-0.00685699	0.00320195	-0.317891	-0.204228	-0.0148839	1	0.76374
std_ctr	-0.0293729	-0.0777166	0.20158	-0.093876	0.236835	-0.0163799	0.043057	-0.296883	-0.152919	-0.13145	0.76374	1
rating	0.0441757	0.0187324	0.330144	-0.0112216	0.253739	0.00360718	0.233578	-0.0203942	0.113816	-0.0840605	0.0234906	0.0256019

Roberto Noel
Gus Xia
Independent Study: MIR
11/12/19

From clustering based on these highly correlated features we gain some insights:





	std_cent	std_rolloff	std_zcross	std_flt
0	411.067219	863.785522	0.120216	0.010395
1	259.688641	441.822822	0.096759	0.014422
2	125.725767	165.870927	0.044023	0.016835

As we can see cluster two is the lowest rated of the them all, with the lowest std_cent, std_rolloff and std_zcross. These are all measures of changes in energy at different frequency levels, so a low rating when these are less present makes sense. It is likely that there are many flat tones within this group which would logically bring ratings down. A highly melodic song would not be included in this cluster for example.

In the next progress report, I will do clustering based on tree features which will hopefully give us similar insights. I'm currently working on functions to extract these features, which will also likely improve the results of the neural network. I am also working to improve the extraction of tempo from these songs, since I have noticed some inaccuracies in librosa's extraction. I think tempo will end up being much more correlated with ratings if it is extracted properly.