Overview

Motivation
- Helps online music companies such as Spotify or Apple Music to manage their music base.
- Saves time and effort from manual classification.

Objectives
- Build deep learning (C-RNN) models to automatically classify music genres for real time.
- Improve baseline models accuracy by C-RNN.

Free Music Archive
- The dataset contains 8000 tracks of 30 seconds clips, with 8 balanced genres[^2] listed below.

<table>
<thead>
<tr>
<th>Instrumental</th>
<th>Rock</th>
<th>Electronic</th>
</tr>
</thead>
<tbody>
<tr>
<td>International</td>
<td>Pop</td>
<td>Experimental</td>
</tr>
<tr>
<td>Hip-hop</td>
<td>Folk</td>
<td></td>
</tr>
</tbody>
</table>

70% training / 20% Validation / 10% Test

Dataset

Results & Discussion

Results

<table>
<thead>
<tr>
<th>Models</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random guess</td>
<td>12.5%</td>
</tr>
<tr>
<td>K nearest neighbors</td>
<td>36.38%</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>42.25%</td>
</tr>
<tr>
<td>Multilayer perceptron</td>
<td>44.88%</td>
</tr>
<tr>
<td>Support vector machine</td>
<td>46.38%</td>
</tr>
<tr>
<td>C-RNN</td>
<td>65.32%</td>
</tr>
</tbody>
</table>

Discussion
- Genre “Experimental” is hard to be classified correctly.
- Classifications of other genres perform well.
- CONV layers extracted useful genre clips, listen to our demos.
- Recurrent models enable us to do real-time classification.
- C-RNN does not include music metadata, while baseline model does.

Architecture
- Input: mel-spectrogram
- 3 Convolutional layers
  - Batch Normalization
  - ReLU activation
  - Dropout Regularization
- Recurrent layers
- Output: probability of each genre.

Future
- Understand why C-RNN cannot perform well on “Experimental” genre and improve the accuracy of that genre.
- Consider adding music metadata to C-RNN models and further improve the accuracy.
- Implement an user interface to allow users input a music clip and visualize the real-time music classification online.

Reference