Salsa Subgenre Analysis via Generative Latent Spaces

Attila Barna & Balázs Menkó

Eötvös Loránd University

Advanced Machine Learning Laboratory
March 21, 2025

Project Overview: Salsa Subgenre Analysis

- **Problem:** Understanding the characteristics and relationships between different salsa subgenres (e.g., Cuban Son, Salsa con Rumba, Timba, Salsaton, Mambo).
- **Goal:** Explore and visualize how a generative model's latent space represents and distinguishes between salsa subgenres.
- **Why it's interesting:** Leverages generative AI for data exploration and knowledge discovery in music information retrieval. Provides insights into the nuances of salsa music.
- **Generative AI Component:** Train a Variational Autoencoder (VAE) and analyze its latent space to reveal subgenre similarities and differences.

Project Overview: Salsa Subgenre Analysis

Popular Styles of Salsa Dances

Roadmap

- **Data Collection & Preparation**
  - Find or create a dataset of salsa tracks with subgenre labels (e.g., using Spotify API, online music resources).
  - Initial data exploration and cleaning. Focus on a manageable number of subgenres (3-4 to start).
  - Split data into training, validation, and test sets.
- **Feature Extraction**
  - Extract audio features (e.g., MFCCs, Chroma features, Spectral Contrast, Tempo) using Librosa or similar libraries.
  - Experiment with different feature combinations to find the most informative set.
- **VAE Model Selection & Implementation**
  - Design and implement a Variational Autoencoder (VAE) architecture. Keep the architecture relatively simple.
  - Choose appropriate hyperparameters.

Roadmap

- **VAE Training**
  - Train the VAE on the extracted audio features from all subgenres.
  - Monitor training loss and validation loss to ensure the VAE is learning properly.
- **Latent Space Exploration**
  - Encode the entire dataset into the trained VAE's latent space.
  - Visualize the latent space using dimensionality reduction techniques (t-SNE, PCA) and color-code by subgenre.
- **Analysis & Reporting**
  - Analyze and interpret the latent space visualization:
    - Are the subgenres well-separated in the latent space?
    - Are there any overlaps or unexpected clusters?
    - What might these relationships tell us about the musical characteristics of the subgenres?
  - Prepare a final presentation summarizing the project, results, and insights. Discuss limitations and possible future directions.

Tools and Technologies

- **Programming Language:** Python
- **Audio Processing:** Librosa, PyAudioAnalysis
- **Machine Learning:** TensorFlow/Keras or PyTorch
- **Data Handling:** Pandas, NumPy
- **Dimensionality Reduction:** Scikit-learn
- **Visualization:** Matplotlib, Seaborn
- **IDE:** VS Code/Cursor
- **Version Control:** GitHub
- **Cloud Computing:** Google Colab (for GPU access)

Generative AI Component: VAE Latent Space Analysis

- **Variational Autoencoder (VAE):** A generative model that learns a probabilistic representation (latent space) of the input data.
- **Latent Space Encoding:** Projecting audio feature vectors into the VAE's latent space provides a compact, abstract representation of each track.
- **Visualization & Interpretation:** Visualizing the latent space, color-coded by subgenre, reveals patterns and relationships between subgenres based on their underlying audio features. Areas where points of the same color are grouped suggests the model has learned to extract similar features in each subgenre.

Expected Outcomes

- **Visual Representation of Subgenre Relationships:** A clear visualization of how different salsa subgenres are related based on their audio features.
- **Insights into Musical Characteristics:** Identification of the audio features that distinguish between the subgenres, as revealed by the VAE's latent space representation.
- **Understanding of VAEs for Music Analysis:** Practical experience training and interpreting a VAE for music data.

Potential Next Steps (Beyond This Project)

- Train the VAE on raw audio waveforms or spectrograms directly for improved feature extraction.
- Use the latent space for conditional music generation of salsa subgenres.
- Compare different generative models (LSTM etc.) for salsa subgenre analysis.