DL4DS Final Project Proposal: *Music Genre Classification*

Matt Maslow & Jon Neimann

November 4th, 2024

Abstract

This project aims to develop a music genre classification system. It will compare a K-Nearest Neighbors (KNN) approach with a Convolutional Neural Network (CNN) to explore the strengths and limitations of traditional versus deep learning methods. By leveraging audio feature extraction and spectrogram-based representations, we will assess the effectiveness of each approach in genre recognition, evaluating their accuracy and adaptability to diverse musical genres.

Introduction

As music streaming services are popular, genre classification systems are integral to applications like recommendation engines and digital libraries. However, blending musical genres and subjective genre boundaries poses challenges for traditional classification methods. In this project, we propose a hybrid approach, combining a simple yet effective KNN classifier with a deep learning-based CNN model. The CNN will analyze spectrograms, allowing us to capture intricate patterns within audio data that simpler algorithms might miss. By comparing these approaches, we aim to understand how deep learning can improve genre classification accuracy and handle complex audio data better.

Related Work

Previous music genre classification work has explored traditional and deep learning methods. CNNs have demonstrated strong performance in audio tasks by effectively extracting features from spectrograms, which can represent detailed temporal and frequency information. Building on these findings, we will examine how KNN and CNN models perform in genre classification. We will provide insights into the trade-offs between traditional machine learning and deep learning for music analysis applications. This approach aligns with the real-world capabilities of platforms like Shazam and voice assistants that rely on accurate music categorization for user interaction.

Proposed Work

Our methodology will include implementing both a KNN classifier and a CNN for genre classification using extracted audio features and spectrograms:

1. *Feature Extraction and Spectrogram Generation*:
   - Extract audio features such as Mel-frequency cepstral coefficients (MFCCs), spectral contrast, and tonnetz.
   - Generate spectrograms of each audio file, transforming audio signals into visual representations that will be the input for the CNN.
2. *K-Nearest Neighbors Classifier*:
   - Train a KNN model on the extracted audio features.
   - Evaluate its performance across different distance metrics (e.g., Euclidean, Manhattan).
3. *Convolutional Neural Network Model*:

Implement a CNN architecture designed for spectrogram classification. This will allow the model to learn complex patterns within audio data that are specific to genre characteristics.
- Experiment with different architectures and hyperparameters to optimize CNN performance.

4. *Comparison and Evaluation*:
- Assess the performance of both KNN and CNN models on genre classification tasks.
- Evaluate the models on metrics such as accuracy, precision, and recall, highlighting the comparative strengths of each approach in handling high-dimensional audio data.

5. *Hyperparameter Tuning and Cross-Validation*:
- Conduct hyperparameter optimization for KNN and CNN models, using grid search and cross-validation to ensure robust performance.

Datasets

We will use the GTZAN dataset, which contains 1,000 audio tracks across ten genres, for primary training and evaluation. For the deep learning model, we may explore larger datasets such as the Million Song Dataset or FMA (Free Music Archive)** to enhance model generalization, given that CNNs benefit from greater data variety.

Evaluation

Our evaluation will include:
- *Confusion Matrix Analysis*: To identify common genre misclassifications.
- Precision, Recall, and F1-Score: For each genre, providing detailed insights into model performance.
- *Model Comparison*: We will highlight areas where CNN outperforms KNN and vice versa, examining which features are best captured by each method and under what conditions CNN's deep feature extraction provides an edge.

Timeline

- *Weeks 1-2*: Literature review and dataset preparation.
- *Weeks 3-4*: Feature extraction, spectrogram generation, and initial implementations of KNN and CNN models.
- *Week 5*: Model training and hyperparameter tuning for both KNN and CNN.
- *Week 6*: Model evaluation, comparison, and analysis of results.
- *Week 7*: Final adjustments, documentation, and report submission.

References

[1] Keoikantse Mogonediwa, *Music Genre Classification: Training an AI Model*, 2024.
[2] Nithil Mariya Stephen, *Music Genre Classification*, California State University, Northridge, 2023.