Song Lyrics Genre Classification

Project Group 18

Claudia Gentili : 581522
Simone Ianniciello : 581201
Alessandro Stefanelli : 686084

Under the Guidance of
Prof. Lucia PASSARO

Dept. of Computer Science
Università di Pisa

November 17, 2024

Overview

Introduction: Importance of Song Classification

Context

- **Growth**: Exponential increase in song availability on the web.
- **Popularity**: Rise of music streaming platforms.

Significance

- **User Impact**: Essential for enhancing user experience through effective music categorization and recommendation.
- **Engagement**: Classification influences user satisfaction and platform engagement.

The Problem of Genre Classification

Problem

- Songs often *don't fit* into a **single** genre
- **Manual** classification is impractical and subjective

Need

- **Automation**: Necessity for automated music classification in vast and expanding music collections.

Lyrics-Based Approach and Study Objective

Traditionally Approach

Genre classification has relied on various features extracted from music audio signals such as

- melody
- harmony
- rhythm
- timbre

Lyrics-Based Approach

- Song lyrics provide thematic and stylistic information
- Analyzing lyrics with advanced NLP techniques

Study Objective

- Explore genre classification based on lyrics
- Utilize pre-trained language models

Prior Work

Neural Methods

- Costa et al. [1]: Compared CNNs and SVMs using spectrograms.
- Jeong and Lee [2]: Used temporal features in audio with a deep neural network.
- Bahuleyan [3]: Success of combining convolutional and hand-crafted features in ensembles.

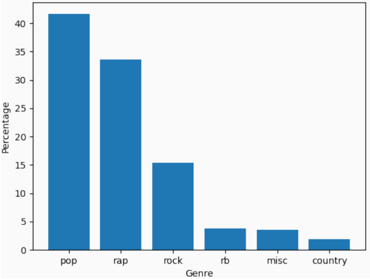
Lyrics-Based Classification

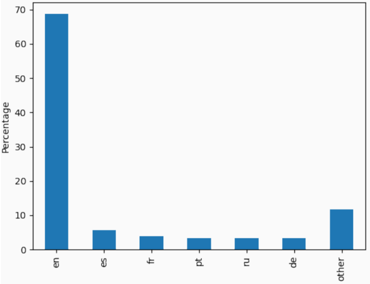
- Mayer et al. [4]: Evaluated models on textual features from song lyrics.
- Lima et al. [5]: Proposed BILSTM for Brazilian song lyrics.
- Tsaptsinos [6]: Used hierarchical attention networks.
- Boonyanit et al. [7]: Developed GloVe embeddings for LSTM models, achieving an accuracy of 68%

Pre-Trained Language Models

- Akalp et al. [8]: Investigated BERT and DistilBERT for genre classification.

Dataset





- Genius Song Lyrics Dataset from Kaggle
- Contains song lyrics up to 2022
- Information includes:
  - title
  - artist
  - year
  - genre
  - views
  - etc...

Data Preprocessing

Steps

1. **Remove Incomplete Entries:** Eliminated entries with 'None' values.
2. **Exclude Non-Musical Pieces:** Removed `misc` tagged entries.
3. **English Texts Only:** Focused on English lyrics.
4. **Retain Popular Tracks:** Kept tracks with over 1,000 views for reliability.
5. **Balance by Tag Attribute:** Subsampled to balance tags and avoid genre overrepresentation.
6. **Remove Metadata:** Content removed in square brackets.

Result

A tag-balanced dataset of 70,178 entries.

Models Overview

We focus on the **performance** of Baseline models, RNNs, and pre-trained language models under **partial** and **complete** fine-tuning, comparing F1 score, precision, and recall

Baseline Models

Feedforward neural networks with tokenization and FastText embeddings

Recurrent Neural Models

RNNs with GloVe embeddings

Pre-trained Models

DistilRoBERTa, DistilGPT2

Baseline Model

Model

Feedforward Neural Network

Model Architecture

- **Hidden Layer:** Single layer with 300 units.
- **Batch Normalization:** Utilizes ReLU activation function.

Preprocessing Steps

1. **Tokenization:** Using NLTK word tokenizer.
2. **Non-Alphabetic Tokens Removal:** Eliminates tokens without alphabetic characters.
3. **Punctuation Removal:** Removes all punctuation.
4. **Stopwords Removal:** Excludes common stopwords.
5. **Conversion to Lowercase:** Converts all text to lowercase.
6. **Embedding Calculation:** Computes 300-dimensional FastText embeddings.
7. **Vector Averaging:** Takes the mean of these embedding vectors.

Pre-trained Models

Recurrent Neural Models

- **Approach**:
  - Uses Recurrent Neural Networks to scan lyrics.
  - Last token's context used as song embedding for linear classifier.
- **Embedding**: 300-dimensional token embeddings with GloVe.
- **Recurrent Layers**: Based on GRU units.
- **Context Length**: Limited to 150 tokens due to parallelization issues.

Pretrained Language Models

DistilRoBERTa:

- *Discriminative* model with 82.8M parameters.
- Pretrained with bidirectional context.

DistilGPT2:

- *Generative* model with 88.2M parameters.
- Causally pretrained model.

Experiments

We began training and evaluating the model to maximize accuracy on new data sets.

Training Process

- **Epochs**: Up to 1000, with early stopping based on validation loss.
- **Optimization**: Cross entropy loss and Adam optimizer.
- **Dataset Split**: 60% training, 20% validation, 20% testing.

Model Selection

- **Validation**: Use of hold-out validation set for selection.
- **Criteria**: Best model based on macro F1 score.
- **Retraining**: Final retraining on combined training and validation set.

Additional Details

RNN and Pretrained Models

- Use Optuna for hyperparameter search.
- 3 epochs for both initial and final retraining phases.

Summary of Model Performances

Model	**Fine-Tuning**	**Macro-Averaged F1 Score**	**Best Genres**
Baseline Model	None	0.55	Country, Rap
RNN	None	0.55	-
DistilGPT2	Partial	0.19	-
DistilGPT2	Full	0.66	-
DistilRoBERTa	Partial	0.59	Rap
DistilRoBERTa	Full	0.66	Rap, Country

Error Analysis

Observations

- **Misclassifications**: Frequent errors between similar genres, such as pop and rock.
- **Recognition Issues**: The least well-recognized genre is always pop, highlighting its mixed nature.

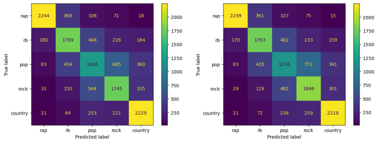


Figure: Left: DistilGPT2 (full), Right: DistilRoBERTa (full)

Insights

- **Confusion Matrices**: Evidence from confusion matrices.
- **Cause**: Blending of stylistic elements in lyrics confuses models.

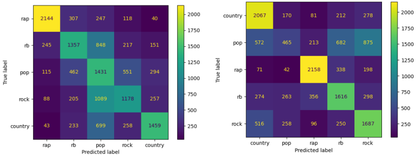


Figure: Left: Baseline model, Right: RNN model

Conclusions

Our Conclusions

- Pre-trained models show good performance in genre classification
- Full fine-tuning significantly improves results
- Achieved notable average F1 score of 0.66 in complete fine-tuning scenarios.

Future Research Directions

1. **Integrate Audio Features**: Enhance genre classification using melody and rhythm.
2. **Model Comparison**: Evaluate trade-offs by comparing distilled models with larger versions.
3. **Enhance Optimization**: Implement advanced regularization and data augmentation.

References

Yandre M.G. Costa, Luiz S. Oliveira, Carlos N. Silla (2017)
An evaluation of Convolutional Neural Networks for music classification using spectrograms
Applied Soft Computing 52, 28-38.
https://www.sciencedirect.com/science/article/pii/S1568494616306421

Il-Young Jeong, Kyogu Lee (2016)
Learning temporal features using a deep neural network and its application to music genre classification
In Proceedings of the *International Conference on Pattern Recognition*.

Hareesh Bahuleyan (2018)
Music Genre Classification using Machine Learning Techniques
arXiv preprint arXiv:1804.01149

Maximilian Mayerl, Michael Vötter, Manfred Moosleitner, Eva Zangerle (2020)
Comparing Lyrics Features for Genre Recognition
In Proceedings of the 1st Workshop on NLP for Music and Audio (NLP4MusA), 73-77.

Raul de Araújo Lima, Rômulo César Costa de Sousa, Simone Diniz Junqueira Barbosa, Hélio Cortês Vieira Lopes (2020)
Brazilian Lyrics-Based Music Genre Classification Using a BLSTM Network
arXiv preprint arXiv:2003.05377
https://arxiv.org/abs/2003.05377

Alexandros Tsaptsinos (2017)
Lyrics-Based Music Genre Classification Using a Hierarchical Attention Network
arXiv preprint arXiv:1707.04678
https://arxiv.org/abs/1707.04678

Stanford CS224N, Megan Leszczynski, Anna Boonyanit, Andrea Dahl (2021)
Music Genre Classification using Song Lyrics
https://api.semanticscholar.org/CorpusID:235344286

Hasan Akalp, Enes Furkan Cigdem, Seyma Yilmaz, Necva Bölücü, Burcu Can (2021)
Language Representation Models for Music Genre Classification Using Lyrics
In Proceedings of the 2021 International Symposium on Electrical, Electronics and Information Engineering
https://api.semanticscholar.org/CorpusID:236145441

The End