Heterogeneous Graph Neural Network for Music Emotion Recognition

Angelo Cesar Mendes da SILVA[1], Diego Furtado SILVA[2], Ricardo Marcondes MARCACINI[1]

[1]Universidade de São Paulo, São Carlos, Brazil
[2]Universidade Federal de São Carlos, São Carlos, Brazil

ISMIR 2022 BENGALURU

Introduction

- Music is intrinsically connected to human emotions
- Emotion labels can be identified from arousal and valence domains
- Musical data have a multimodal and heterogeneous composition
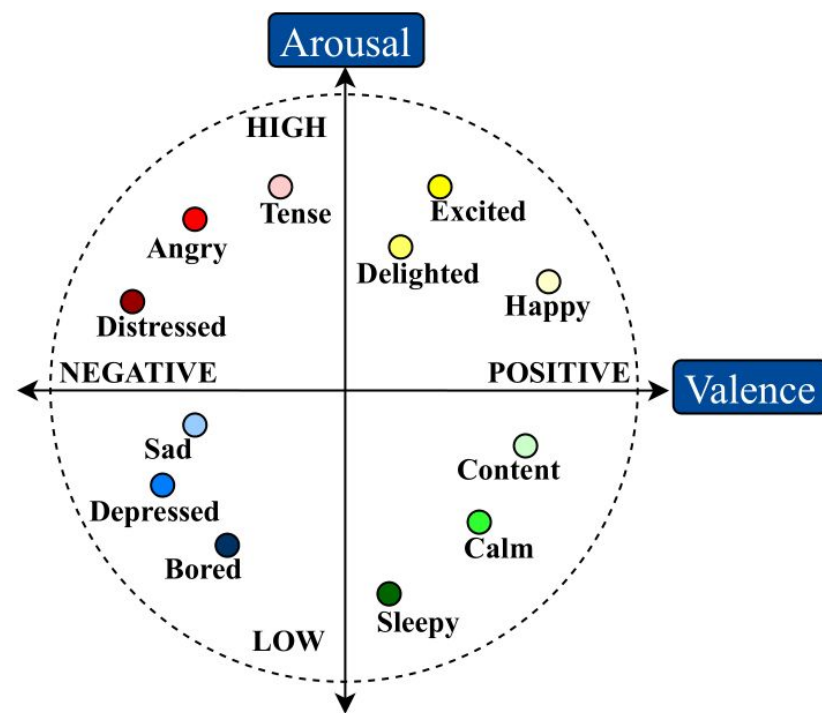
Example of emotions mapped in the arousal and valence domains
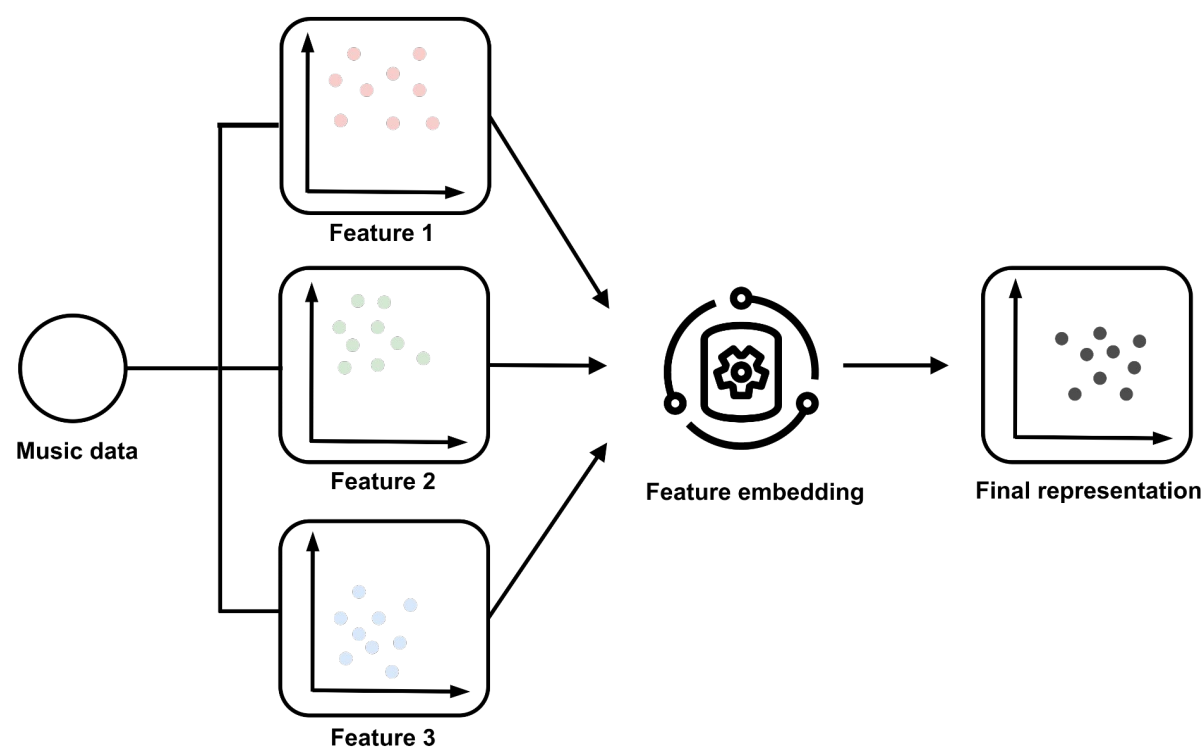
Embedding a music representation from heterogeneous features

Major Contributions

- We structure audio and lyrics features on heterogeneous network
- We applied a network regularization framework to propagate information between nodes
- Our approach learn a new heterogeneous multimodal graph-based music representation
- We evaluated multiple strategies to build a heterogeneous network for musical data.

Proposed Approach
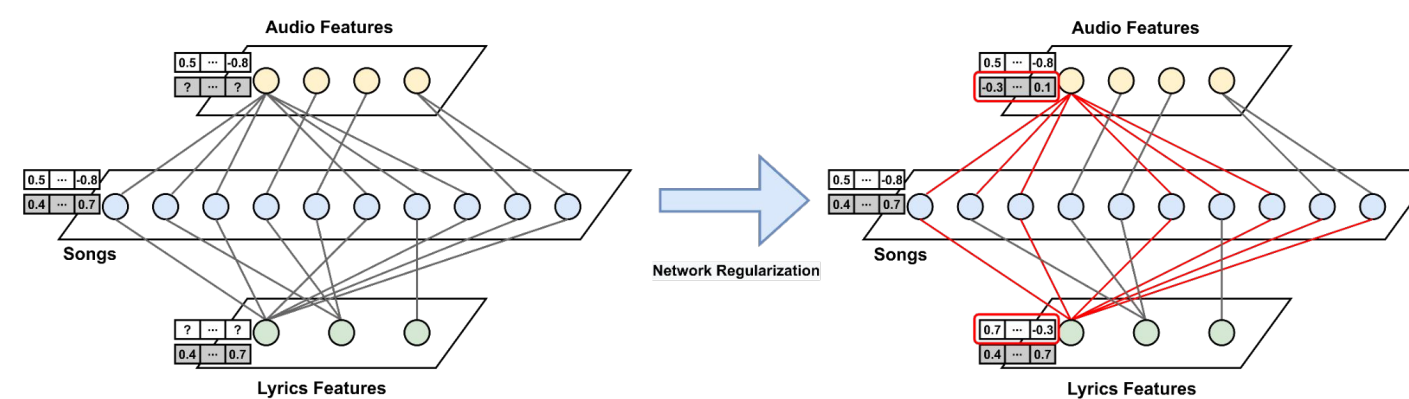
Heterogeneous network regularization to propagate features between nodes
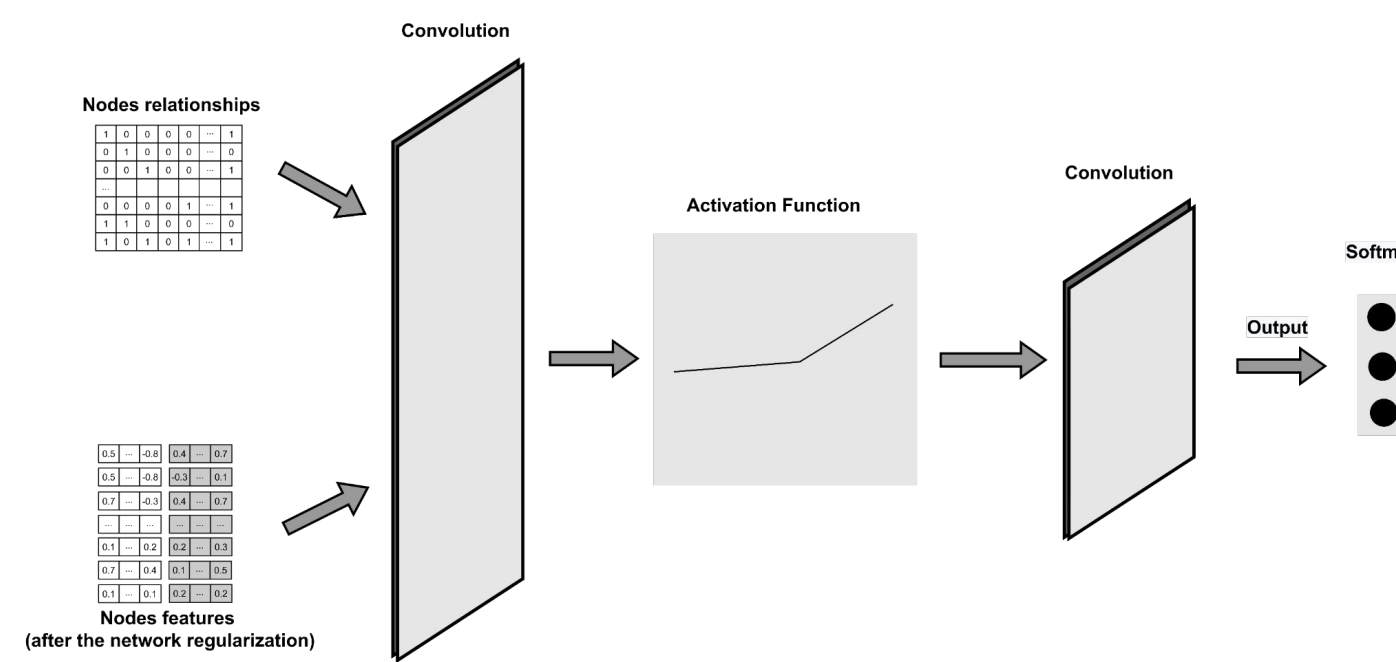
The GCN model uses relationships and feature nodes as input to handle MER.

Experimental Evaluation

Dataset

Task:	Music Emotion Recognition
Songs:	764
Domains:	Arousal and Valence
Classes:	Positive, Negative and Neutral
Features:	Audio and Lyrics

Evaluated representations

Proposed representation
multimodal graph-based representation

Baseline

BoW	BoW + Audio
Bert	Bert + Audio
Audio	

Sampling

Train
Test

Algorithms

Inductive scenario
Multilayer perceptron

Transductive scenario
Label Propagation Algorithm

Results Overview
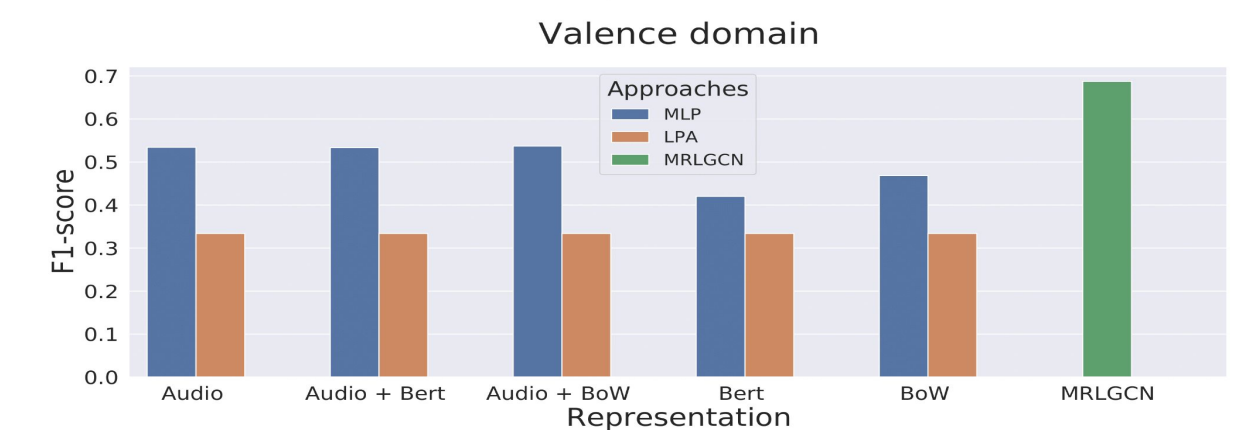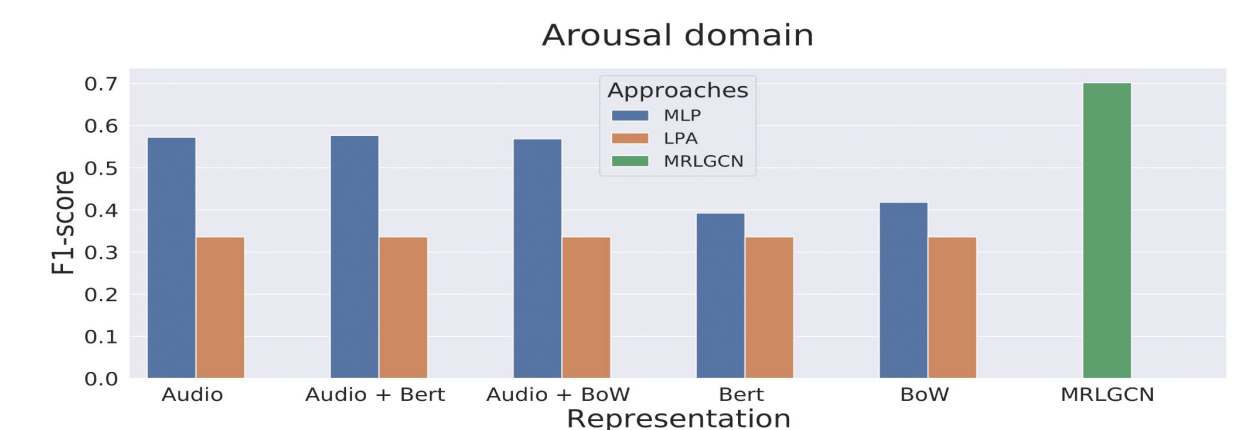
Conclusion

- Our results showed that the proposed representation is more discriminative than the other unimodal or multimodal representations
- The results reinforce the applicability of graph-based representations in unstructured and multimodal data