

Multi-modal User Adaptation for Automatic Music Tagging

Yannis Vasilakis

2022

- Research on music descriptors has been hypothesizing the existence of objective, universal musical properties
- But, each user might have a different definition for "Pop". What is yours?
- Can we measure/represent this "personal" opinion? Are users consistent with their definition?

This is not a Diva House song!

Yup, it's more of a bubblegum dance house song..

But they both sound like doom, tss, doom.

01 Background

Descriptors are inherently subjective and depend on:

- Culture
- Education
- Experience

Machine Learning can be used to extract embedded spaces where every song is represented

Each user will have a different distribution

Metalheads

Scorpions
Guns N' Roses

Iron Maiden
Bruce Dickinson
Anthrax
Metallica
Slayer
Helloween
Blind Guardian
Gamma Ray
Edguy

Mortals

Bruce Dickinson
Anthrax
Metallica
Slayer
Scorpions
Guns N' Roses
Iron Maiden
Edguy
Blind Guardian
Gamma-Ray
Helloween

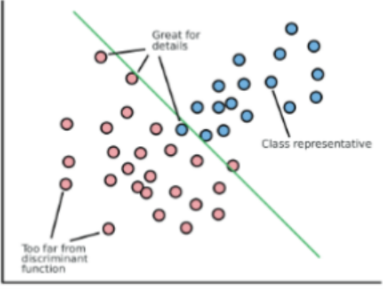
02 Technology

Training from scratch for each user is not computationally tractable

- Users are not available for 40 hours
- Not every data point is informative

What if we could search and query only the informative data points?

- Active Learning to the rescue!



03 Research Questions

- Can we evaluate the personalisation of definitions?
 - Are users consistent with themselves?
- Can personally defined descriptors provide better:
 - Accuracy (through embeddings)
 - Profiling insights (user clustering)
- Are continuous defined descriptors useful?
 - Can we measure "pop-ness"?
 - Is this modality more useful?

04 Future Work

- Hypothesis evaluation
 - Questionnaire to evaluate users self-consistency with their descriptors
- Dataset formation
 - Gather users, information about them, their personal descriptors
- Objective vs Subjective descriptors
 - Find SOTA models for Genre Classification-Tags
 - Compare their personalised accuracy

Finding

Nothing

Question

What is the least amount of song annotations needed to define a tag for a specific user? Do tags (or music descriptors in general) mean the same thing for different users?