# Leveraging the structure of musical preference in content-aware music recommendation


**Paul Magron**, Cédric Févotte

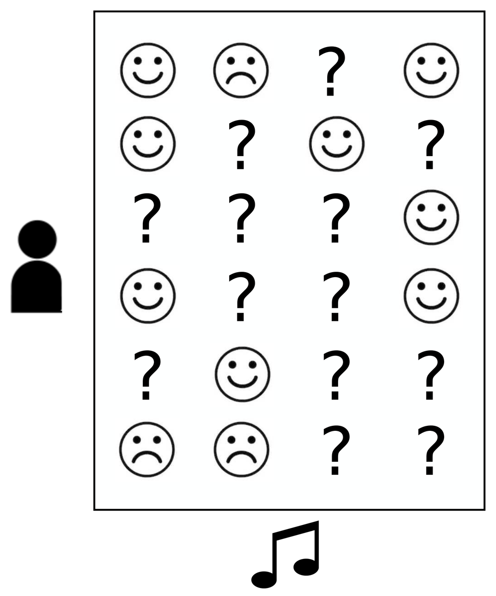
CNRS, IRIT, Université de Toulouse, France



ICASSP2021

Paper ID: 1629


## Music recommendation



- A task at the core of commercial music streaming platforms (Spotify, Deezer...).
- Predict a set of items (=songs) that a user might be interested in.
- Main idea: exploiting *similarities* between users and/or items.

### The cold-start issue [1]

New songs recently added to the catalogue do not have any listening history: recommendation cannot be based solely on users' similarities.

## Main techniques

Collaborative filtering: Users with similar tastes in the past (collected data) will have similar tastes in the future (predictions).
Content-based-approaches: Users who liked some songs (collected data) will like songs with a similar content (predictions).
Content-aware recommendation: incorporate content as side information in collaborative filtering.

| | Performance | Cold-start |
|---|---|---|
| Collaborative filtering | ✓ | ✗ |
| Content-based | ✗ | ✓ |
| Content-aware filtering | ✓ | ✓ |

### Problem

Usual content features do not have a clear music-related meaning: they might not be optimal for recommendation.

## Content-aware collaborative filtering

### Data

- $\mathbf{Y} = \{y_{u,i}\} \in \mathbb{R}_+^{U \times I}$ = users / items interactions.
- For implicit feedback (here: playcounts) *binarized* data $\mathbf{R}$ is preferred.

### Content-aware weighted matrix factorization [2]

$$\mathbf{R} \approx \mathbf{W}^\mathsf{T}\mathbf{H} \text{ with } \mathbf{h}_i \approx \phi(\mathbf{z}_i)$$

- $\mathbf{W} \in \mathbb{R}^{K \times U}$: users preferences.
- $\mathbf{H} \in \mathbb{R}^{K \times I}$: songs attributes.
- $\mathbf{z}_i \in \mathbb{R}^L$ : content vector extracted from low-level acoustic features $\mathbf{x}_i$.
- $\phi : \mathbb{R}^L \to \mathbb{R}^K$ mapping between content and item attributes. Here, linear mapping $\phi(\mathbf{z}_i) = \mathbf{B}\mathbf{z}_i$

Estimation:

$$\min_{\mathbf{W},\mathbf{H},\mathbf{B}} \sum_{u,i} c_{u,i}(r_{u,i} - \mathbf{w}_u^\mathsf{T}\mathbf{h}_i)^2 + \lambda_W \sum_u ||\mathbf{w}_u||^2 + \lambda_H \sum_i ||\mathbf{h}_i - \mathbf{B}\mathbf{z}_i||^2$$

- $c_{u,i}$ is a confidence weight computed from the raw playcount $y_{u,i}$.
- Alternating least squares: iterative estimation procedure with closed-form updates.

### Proposed content features

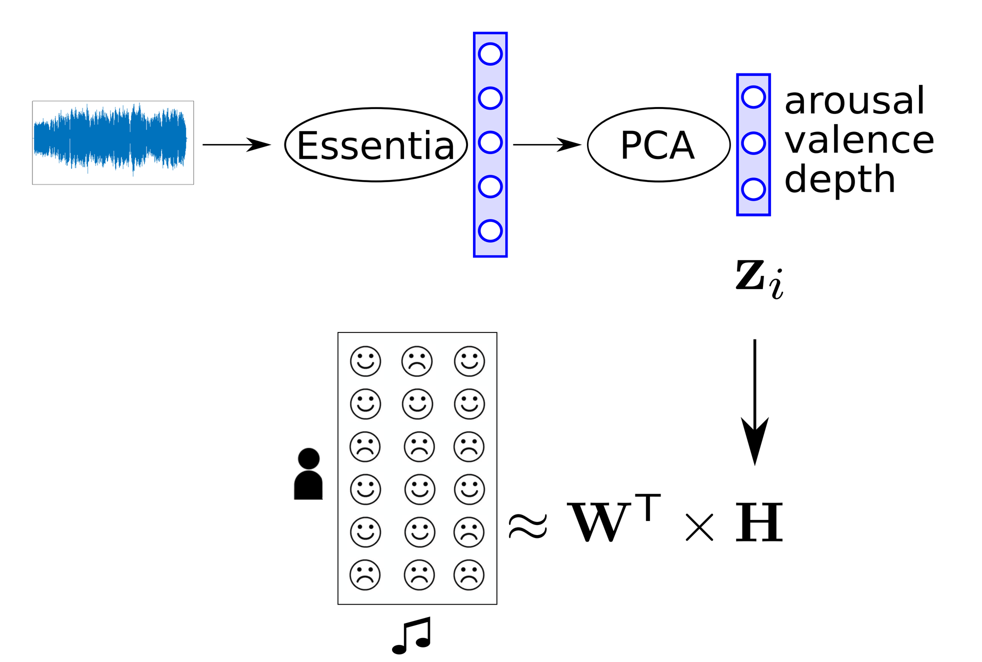
- Research in music psychology [2] shows that musical preference can be described using 3 factors:
  - *Arousal*: is it energetic/intense or calm?
  - *Valence*: is it sad or happy?
  - *Depth*: is it "sophisticated" or simple?
- Computing the AVD model:
  - Collect a set of high-level features from acoustic descriptors $\mathbf{x}_i$ using the Essentia toolbox.
  - Perform a PCA ($L = 3$) with oblimin rotation.
- Correlations with high-level features:

| | Arousal | Valence | Depth |
|---|---|---|---|
| Rousing/Passionate | **0.56** | -0.11 | -0.12 |
| Fun/Cheerful | -0.03 | **0.78** | 0.01 |
| Humorous/Witty | -0.05 | **0.63** | 0.12 |
| Aggressive | **0.63** | -0.48 | -0.02 |
| Happy | **0.52** | 0.37 | -0.36 |
| Party | **0.69** | 0.05 | 0.40 |
| Relaxed | **-0.84** | 0.01 | 0.14 |
| Sad | **-0.80** | 0.07 | -0.21 |
| Acoustic | **-0.78** | 0.04 | -0.25 |
| Average loudness | **0.59** | 0.14 | -0.07 |
| Danceable | 0.23 | 0.42 | **0.52** |
| Dissonance | **0.86** | -0.03 | -0.04 |
| Dynamic complexity | **-0.57** | 0.07 | 0.21 |
| Electronic | 0.08 | -0.01 | **0.74** |
| Instrumental | -0.35 | -0.06 | 0.23 |
| Tonal | 0.04 | 0.15 | **-0.60** |

## Method

### Training

- Extract the AVD factors for all song.
- Incorporate it as content feature $\mathbf{z}_i$ in weighted matrix factorization.
- Train the model to learn $\mathbf{W}$ and $\mathbf{B}$ (and $\mathbf{H}$).




### Testing (for cold-start recommendation)

- For a novel song, extract its AVD factors $\mathbf{z}_i$.
- Perform predictions through: $\hat{r}_{u,i} = \mathbf{w}_u^\mathsf{T}\mathbf{B}\mathbf{z}_i$.

## Experiments

Data from the Million Song dataset:

- Songs whose Essentia features are available.
- Filter out inactive user/songs.

| | |
|---|---|
| # users | 9,132 |
| # songs | 7,674 |
| # playcounts | 247,414 |
| % playcounts | 0.35 |

### Tasks

In-matrix prediction = traditional recommendation.

- Keep songs (95 %) for which some listening history (interactions with users) is available.
- Split the playcounts:
  - 70 % for training the model.
  - 20 % for tuning the hyperparameters ($\lambda_W$ and $\lambda_H$).
  - 10 % for in-matrix testing.

Out-of-matrix = cold-start recommendation.

- Predictions on the 5 % remaining songs on which the model is not trained.

Results measured with the normalized discounted cumulative gain ($\in [0, 1]$), higher is better.

| | In-matrix | Out-of-matrix |
|---|---|---|
| Pure collaborative filtering | 0.35 | – |
| Pure content | – | 0.19 |
| Proposed (content-aware) | | |
| Essentia (before PCA) | 0.36 | **0.22** |
| AVD (after PCA) | 0.35 | **0.21** |

- Similar performance for in-matrix recommendation.
- The AVD features allows to address the cold-start problem.
- Content-aware filtering > pure content-based approach for cold-start recommendation.
- Slight performance drop when performing dimensionality reduction.

### Summary

- The AVD model of musical preference is useful for cold-start recommendation.
- A light framework with a compact and and meaningful set of features.

## References


[1] Schedl et al., "Current challenges and visions in music recommender systems research", *International Journal of Multimedia Information Retrieval*, vol. 7, no. 2, pp. 95–116, June 2018.

[2] Liang et al., "Content-aware collaborative music recommendation using pre-trained neural networks", *Proc. ISMIR*, October 2015.

[3] Fricke et al., "Measuring musical preferences from listening behavior: Data from one million people and 200,000 songs", *Psychology of Music*, September 2019.