Multi-objective Hyper-parameter Optimization of Behavioral Song Embeddings

M. Quadrana, A. Larreche-Mouly, M. Mauch
ISMIR 2022 , Apple Inc.

Abstract

Song embeddings are a key component of music recommenders. We study the effect of **hyperparameter optimization** of behavioral song embeddings based on Word2Vec on the tasks of next-song recommendation, false neighbor rejection, and artist and genre clustering.

We show that single-objective optimization has **negative side-effects** on the non-optimized metrics and propose a simple **multi-objective optimization** to mitigate these effects.

Objectives

- Define **metrics** for next-song recommendation, false neighbor rejection, and artist and genre clustering.
- Show that **single-objective optimization** can have negative-side effects on the non-optimised metrics.
- Propose a **multi-objective optimization** to combine recommendation and clustering objectives.
- Study the effects of optimization on different buckets of song popularity.
- Study the effects beyond the tasks being optimized.

Methods

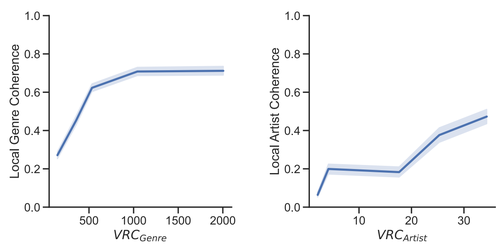
Optimization Tasks

Next-song Prediction: Predict the next-played song given the user history (HitRate and NDCG @100).

False Positive Rejection: Filter out spurious song neighbors, i.e., songs that are neighbors of another by chance (HardNeg @100). Used as safe-guard against optimization side-effects.

Query song	Hard Negative Neighbors
Paradise - Coldplay	Snowman - Sia Immigrant Song - Led Zeppelin Basket Case - Green Day
Smells Like Teen Spirit - Nirvana	Power - Kanye West Ob-La-Di, Ob-La-Da - The Beatles Rock Your Body - Justin Timberlake
Carry On Wayward Son - Kansas	Natural - Imagine Dragons Rap God - Eminen It Ain't Me - Kygo & Selena Gomez

Artist and Genre Clustering: Measure how tracks of the same genre or from the same artist cluster together. Local Genre Coherence positively correlates with Genre Variance Ratio Criterion (same for artists).



Single Objective Optimization

Next-song Prediction: Significant improvements in HitRate/NDCG and reduction in HardNeg. No significant changes in artist and genre clustering.

Artist and Genre Clustering: Large gains in Variance Ratio Criterion, but **regression** in HitRate and NDCG.

Optimizing one metric can hurt the others.

Multi Objective Optimization

Combine next-song prediction and clustering into a single objective with scalarization (shown for genre clustering):

$$\lambda_{Genre}(\alpha) = \alpha \text{HitRate} + (1 - \alpha)\text{rVRC}_{Genre}$$

rVRC_{Genre}: relative change in VRC_{Genre} with respect to baseline
α: relative weight of each objective

MOO discovers configurations that *dominate* the best SOO ones.

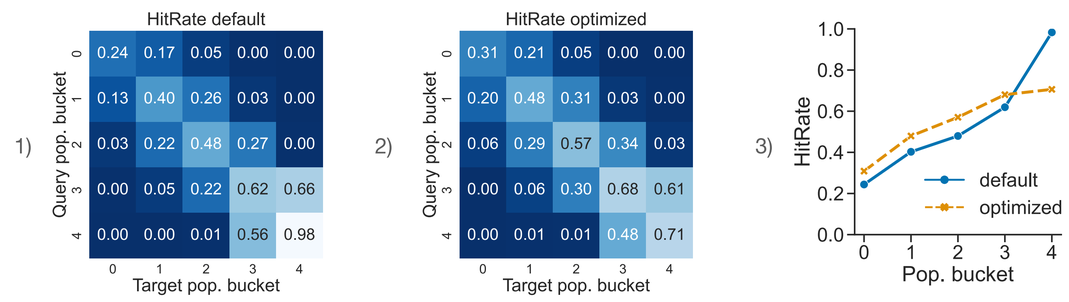
Opt.	Objective	HitRate	HardNeg	VRC_{Genre}	VRC_{Artist}
N/A	N/A	0.3538	0.0128	927	3.373
Single-obj	HitRate	0.3725	0.0108	1033	4.437
	VRC_{Genre}	0.3000	0.0146	**2006**	5.521
Multi-obj	$\lambda_{Genre}(0.01)$	**0.3772**	**0.0084**	1495	5.216

dominates

Results on an internal dataset.

Insights on song popularity

We compute HitRate for song pairs belonging to multiple popularity buckets. Recommendation quality is localized to the nearest buckets to the query and **anti-correlated** with popularity (Fig 1-2).
Optimization **balances** recommendation accuracy across popularity buckets (Fig 3).



Downstream Task

Seeded Radio and Autoplay generate algorithmic streams of songs starting from a user-selected seed. We attempt to predict the first algorithmically played song after the seed (Play Prediction).

On a distinct dataset, we measure the correlation between:
- the cosine similarity of the seed and the first played song
- the ratio of plays (>30s) for that same pair of songs

We observe **stronger correlation** for embeddings optimised with MOO than default and SOO, suggesting that they will perform better in this task.

Conclusions

Multi-objective optimization has substantial benefits over single-objective one:

Finds hyper-parameter configurations **without the negative side-effects** of single-objective optimization.

Balances recommendation accuracy across popularity buckets.

Has potential positive effects on **downstream tasks** such as Play Prediction.