

---

# To Have a Tiger by the Tail: Improving Music Recommendation for International Users

---

Philippe Hamel

Google, Mountain View, CA, USA

HAMELPHI@GOOGLE.COM

## 1. Introduction

At the heart of any music recommendation system is the notion of music similarity. In order to produce relevant recommendations, a system must have some internal representation of how musical entities are related together. Such systems usually learn this representation through different types of data: audio features, user listening data, user ratings, metadata, expert annotations, etc.

One of the most useful signals for building a music similarity representation for recommendation comes from learning a collaborative filtering (CF) space from user data such as user listening behavior and user ratings. However, the distribution of this data is heavily biased towards the most popular music, what we call the head of the distribution. We have much less data on less popular artists from the tail of the distribution. In consequence, the CF space represents relatively accurately head artists, and becomes noisy for tail artists.

This problem becomes even more important when we consider an international user base. Since online services are not uniformly popular or available across the world, the user data distribution is heavier in some countries. Thus, the geopolitical distribution of the users will have a great effect on the CF representation. If this fact is not taken into account, CF models trained over this data will tend to spend most of their capacity modeling popular music from a few countries, while regional or culturally specific music will be somewhat ignored by the model. In practice, when using a naive CF algorithm, locally popular artist representations will tend to cluster tightly in culture-specific clusters. Recommendation models based on these representations will tend to see all artists from small cultural groups as similar, while listeners from this group will consider them as dissimilar.

In this work, we focus on trying to improve the representation of music similarity for tail music content, in particular for content that is specific to a culture or a geographical region, while preserving the quality of the representation for head content. We propose to improve a basic CF model by mixing in external signals and combining local CF models.

## 2. Baseline

### 2.1. Music affinity score

In this work, we focus on the task of finding similar items (track, artist, album, etc.), given a query item in the context of recommendation. The concept of music similarity is ill-defined and depends on what task we are trying to solve. Here, we define similarity in the following way. An item B is similar to item A if B is a good recommendation for users who like item A. What constitutes a good recommendation is also vague and context dependent, but can be empirically measured with offline ground truths or online user feedback. We denote this similarity relationship as an affinity score  $aff(A \rightarrow B)$ . A higher affinity score, means that B is a better recommendation for A. Note that this relationship is not necessarily symmetric, i.e.  $aff(A \rightarrow B) \neq aff(B \rightarrow A)$  in general. To obtain a list of recommendations for an item, we compute the affinity score between the query and all items, and take the top results from the ranked list.

### 2.2. Base model

Given user listening data (play counts, ratings, etc.) it is possible to build a sparse user-item relationship matrix, where items could be tracks, artists or other musically related entities. From this matrix, we can learn a lower dimensional representation of user and items using an SVD-like algorithm. This is what we refer to as the CF space. From the learned representation of the items, we can define a metric to compute an affinity score between two items. For instance, one could define affinity as cosine similarity in the item space. The details of how we obtain this CF affinity is beyond the scope of this work. However, for the sake of argument, let's assume that we can compute a CF-based affinity  $aff_{CF}(A \rightarrow B)$  learned from our user data. This is the base model we aim to improve on.

### 3. Proposed improvements

Starting with the base CF model, how can we improve recommendations for tail content? We first consider mixing in external signals to complement the CF representation. Then, we try to improve the CF representation by combining several local models.

#### 3.1. Mixing affinity signals

One obvious way to improve on the CF based affinity is to mix in signals that do not exhibit the same cultural or popularity bias as the user data with the base CF signal. We can obtain a global affinity score by computing a weighted sum of the signal affinities:

$$aff(A \rightarrow B) = \sum_{i \in \text{signals}} W_i aff_i(A \rightarrow B)$$

where signals include the base CF signal as well as other signals discussed below, and the weights  $W_i$  are optimized through empirical evaluation. The weights can also be a function of the seed query. This could allow, for instance, to put more weight on non-CF signals for queries for which the CF signal is noisy.

##### 3.1.1. CONTENT-BASED FEATURES

One of the most obvious signals to include in a music recommender system is the actual audio signal. Audio features are blind to music popularity and social relationships. They can give a good representation of genre, mood, instrumentation, tempo, etc. This makes them a good complement to the CF signal. One challenge with audio features is to obtain a smooth feature space from which to obtain a relevant affinity score between items. One way to solve this problem is to train a supervised learning model on top of the audio features. In this case, the targets used to train the model represent proxies for audio similarity. Genres, tags or even the CF representation itself can be good targets for training the audio model.

##### 3.1.2. HUMAN-CURATED INFORMATION

One other way to improve recommendations is to use expert annotated data. Annotated data can include information such as genre, instrumentation, mood, melodic content description, etc. Curated data has the advantage of being very reliable. A simple affinity function based on curated data could for instance give a high score when two items exhibit similar relevant characteristics.

The main problem with expert annotated data is that is costly to obtain and hard to scale internationally. It is virtually impossible to have experts cover all of the tail content. Crowdsourcing is a good alternative to expert annotation. It is relatively cheap and easily scalable, at the cost of added noise.

##### 3.1.3. WEB-CRAWLING

It is possible to obtain artist relationship data from other sources on the web. For instance, co-occurrence analysis on Wikipedia or music blogs can provide a good complementary signal for artists for which listening data is sparse or unavailable. Web search data can also be a very useful signal, although it might not be publicly available.

#### 3.2. Training local models

The CF representation can be improved by making a better use of the sparse tail data. One way to do this is to explicitly divide the users into several sub-groups and train CF models on these groups. By dividing the data into smaller targeted datasets, it is possible to better understand the relationship between items within these groups. For instance, to account for the non-uniform geographic distribution, we can train a model for each country. Or, to account for cultural groups within countries, we can divide the data by country and language.

To obtain a global CF affinity function, we can combine the affinity from all the local models:

$$aff_{globalCF}(A \rightarrow B) = \frac{\sum_{i \in \text{datasets}} W_i(A, B) aff_i(A \rightarrow B)}{\sum_{i \in \text{datasets}} W_i(A, B)}$$

where  $W_i(A, B)$  are weights that grow with the popularity of items  $A$  and  $B$  within the dataset  $i$ . This has the effect that groups that know more about items  $A$  and  $B$  will influence more the global affinity for these items.

Dividing the data into explicit user groups requires user demographic information. However, it is still possible to use this technique even without explicit demographic information. If we assume that some user characteristics influence the listening behavior, it should be possible to find these groups from the data. One way to infer user groups is to use an unsupervised clustering algorithm on the user data.

### 4. Conclusion

The task of finding relevant recommendations for tail content is challenging. In this work, we discuss why this is hard, and propose improvements on a base CF model. The solutions proposed include using content-based models, leveraging human-curated data, mixing in external sources, training local models and modeling user characteristics.

### Acknowledgments

This work was done in collaboration with several researchers and engineers. I would like to thank Li Zhang, Rasmus Larsen, Sally Goldamn, Nicolas Boulanger-Lewandowski, Adam Roberts, Pierre Grinspan, Douglas Eck, and Daniel Steinberg for their contributions.