

# Adapting Collaborative Filtering to Personalized Audio Production

Bongjun Kim and Bryan Pardo

EECS Dept., Northwestern University, Evanston, IL 60208  
bongjunkim2013@u.northwestern.edu, pardo@northwestern.edu

## Abstract

Recommending media objects to users typically requires users to rate existing media objects so as to understand their preferences. The number of ratings required to produce good suggestions can be reduced through collaborative filtering. Collaborative filtering is more difficult when prior users have not rated the same set of media objects as the current user or each other. In this work, we describe an approach to applying prior user data in a way that does not require users to rate the same media objects and that does not require imputation (estimation) of prior user ratings of objects they have not rated. This approach is applied to the problem of finding good equalizer settings for music audio and is shown to greatly reduce the number of ratings the current user must make to find a good equalization setting.

## Introduction

Media production tools, such as audio equalizers are widely used in music production and video production. In the past, these tools were mainly used by expert professional engineers. Today, everyone is producing media for web distribution and sharing (e.g. Youtube, Soundcloud, Bandcamp). Thus, the need for media production software a non-expert can use has increased.

Audio equalizers (EQs) are perhaps the most commonly used tools used in audio production. Equalizers selectively boost or cut restricted portions of the frequency spectrum, and in doing so dramatically alter the timbre of a sound. A typical parametric equalizer has on the order of 20 knobs. The large number of controls makes it difficult for non-experts to use effectively.

(Sabin, Rafii, and Pardo 2011) explored an *evaluative interface* for controlling a parametric equalizer. Systems using the evaluative approach ask the user to name a goal (e.g. make the sound “warm”) and then rate (evaluate) a variety of examples in light of that goal (How warm is it when I apply *this* equalization?). Correlation between how

examples vary and the user’s ratings of examples are used to create a personalized effect. For EQ, systems typically require the user rate roughly 25 examples to do so.

Since rating 25 examples can be time-consuming, researchers proposed transfer learning (Pardo, Little, and Gergle 2012) to reduce the number of ratings required. The approach relies on a database of many prior user ratings of examples. *SocialEQ* (Cartwright and Pardo 2013) is a web-based data-collection tool that used the method in (Pardo, Little, and Gergle 2012) to learn the EQ setting associated with hundreds of words (e.g. Bob’s EQ setting for “warm” sound), making it a useful knowledge base.

Unfortunately, the transfer learning method in (Pardo, Little, and Gergle 2012) requires all users to have rated the exact same set of examples so that distance between users can be directly measured by comparing user ratings of examples. *SocialEQ* had users rate a randomly-selected 25 examples out of a set of 50 examples, so any two users only overlap on a portion of their rated examples. This made prior users not directly comparable.

We present a way to apply prior user data, without need for prior users to have all rated the same set of examples. This method can be adapted to any situation where collaborative filtering is desirable, the end products created for users are comparable to each other, but prior users did not rate the same set of examples as the current user.

## The Method

A typical approach to using prior user ratings to inform current learning is collaborative filtering (Jannach et. al 2010). Here, the system provides a suggestion for something the current user may like, but has not yet rated (e.g. a movie on Netflix) by combining the ratings of prior users across objects the current user has not rated. Then, the highest rated object is returned to the current user as a suggestion (e.g. You may like “Shrek”).

The influence of a prior user is weighted by how similarly the current user and that prior user rated some set

of objects. If a prior user has not rated some object that the current user has rated, a rating may be estimated by taking an average value across users or across objects.

In our case, the end result is a new, personalized item (a 40-band EQ curve) not in the set of rated objects. The equalization curves learned for any two users can be directly compared, even though they were generated from different sets of rated objects. We leverage this to our advantage. Instead of filling in estimates for “missing” ratings so that all prior users can be compared, we create an EQ curve from the current user’s set of ratings and compare that curve to the EQ curves learned for each prior user. This lets us apply data from prior users, even if they rated completely different sets of example EQ settings.

The method is as follows: We ask the user to rate a small number of examples (e.g. 5). From this, we build a 40-band EQ curve using the active learning variant from (Pardo, Little, and Gergle 2012). This method requires 25 ratings to produce a good curve, but a curve learned from fewer rated examples is sufficient to locate the user in the space of prior users. We compare this (admittedly bad) EQ curve, using a 40-point Pearson correlation), to each of the EQ curves learned from previous users. We then create a composite EQ curve for the current user from the 64 closest EQ curves from prior users. The weight of each prior users’ EQ curve (learned from 25 examples) is proportional to its similarity to the current user’s curve.

## The Experiment

We measured the effectiveness of this approach on the *SocialEQ* data set. A user-concept is a 40-band EQ curve learned from user ratings of 25 equalization settings. We used the 1635 user-concepts deemed to be from reliable contributors (see (Cartwright and Pardo 2013) for how they measure reliability). The baseline method generates an EQ curve from an individual’s ratings to randomly selected examples, without any user of prior data. The test method was the one described in the prior section.

We compared methods as follows: Select a user from the data and treat it as the “new” user. Build the EQ curve from N rated examples using the active learning. Then, build the curve using our new method. Since we know the final EQ curve learned from 25 ratings (the user-concept) we can compare both new EQ curves to the user-concept. The EQ curve most similar to the user concept is the best. We use Pearson’s similarity measure between the learned EQ curve and the user-concept curve as the similarity measure. Values range from 1 (perfect correlation) to -1 (inverse correlation). Higher numbers are better.

For each value of N we repeat this for all 1635 user-concepts and take the average value for each condition (“baseline” and “using prior data”). We then plot this

average value as a function of the number of rated examples used to construct the new EQ curve. Fig. 1 shows, the new method (red solid line) beats the baseline (dashed line) if users rate fewer than 20 examples.

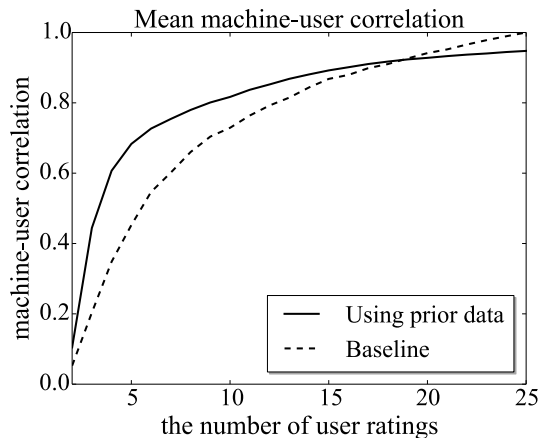


Figure 1. Learning effectiveness as a function of the number of ratings.

## Conclusions

We presented a way to use prior user data to reduce the number of ratings required for creating a personalized audio EQ curve. Our method does not need users to rate the same set of objects and does not need data imputation to fill in missing ratings. The experiment shows that comparable results to the baseline can be achieved using half as many (e.g. 7 vs. 14) rated examples. This approach is generalizable to any case where the final output is a personalized item that can be placed in a metric space.

## Acknowledgements

This research was supported by grant 1116384 from the US National Science Foundation.

## References

- Cartwright, M. and Pardo, B. 2013 Social-eq: Crowdsourcing an equalization descriptor map. In *Proceeding of International Society for Music Information Retrieval* 395-400, Curitiba, Brazil.
- Jannach, D., Zanker, M., Felfernig, A., and Friedrich, G. 2010. *Recommender Systems: An Introduction.* Cambridge
- Pardo, B., Little, D., and Gergle, D. 2012. Building a personalized audio equalizer interface with transfer learning and active learning. In *Proceedings of the second international ACM workshop on Music information retrieval with user-centered and multimodal strategies, 13-18*, New York, NY, USA: ACM
- Sabin, A. T., Rafii, Z. and Pardo, B. 2011. Weighting-Function-Based Rapid Mapping of Descriptors to Audio Processing Parameters. *Journal of the Audio Engineering Society* 59: 419-430.