

Beyond Globally Optimal: Focused Learning for Improved Recommendations

Alex Beutel^{1*}, Ed H. Chi¹, Zhiyuan Cheng^{2†}, Hubert Pham¹, John Anderson¹

¹Google, Inc. ²Pinterest

{alexbeutel, edchi, hubertpham, janders}@google.com

ABSTRACT

When building a recommender system, how can we ensure that *all* items are modeled well? Classically, recommender systems are built, optimized, and tuned to improve a global prediction objective, such as root mean squared error. However, as we demonstrate, these recommender systems often leave many items badly-modeled and thus under-served. Further, we give both empirical and theoretical evidence that no single matrix factorization, under current state-of-the-art methods, gives optimal results for each item.

As a result, we ask: how can we learn additional models to improve the recommendation quality for a specified subset of items? We offer a new technique called *focused learning*, based on hyperparameter optimization and a customized matrix factorization objective. Applying focused learning on top of weighted matrix factorization, factorization machines, and LLORMA, we demonstrate prediction accuracy improvements on multiple datasets. For instance, on MovieLens we achieve as much as a 17% improvement in prediction accuracy for niche movies, cold-start items, and even the most badly-modeled items in the original model.

Keywords

recommendation; regularization

1. INTRODUCTION

How can we predict what movies a user will like? Or to which users an app would be appealing? How can we ensure that all movies or apps have a good opportunity to be surfaced? Recommender systems have become an integral part of our everyday lives, from Netflix recommending movies to Yelp suggesting restaurants to Google Play offering music and apps. While these uses of recommender systems have clearly been successful, little research has focused on ensuring that *all* items in these recommender systems are modeled

* A portion of this work was done while at Carnegie Mellon University.

† A portion of this work was done while at Google, Inc.

©2017 International World Wide Web Conference Committee (IW3C2), published under Creative Commons CC-BY-NC-ND 2.0 License.
WWW 2017, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4913-0/17/04.
<http://dx.doi.org/10.1145/3038912.3052713>

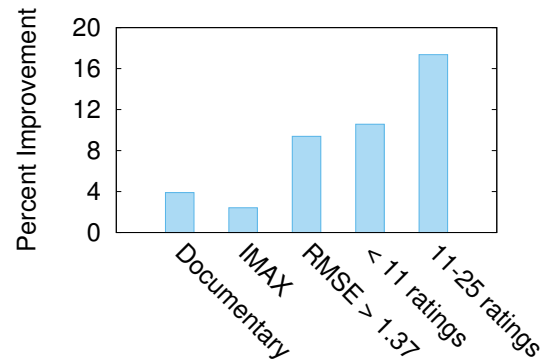


Figure 1: Focused learning improves MovieLens predictions for under-served categories of movies, including (1) niche genres, (2) items for which we have few observations, and (3) even the most badly-modeled items from our original model.

well. For these systems to be continually trusted, it is crucial that we understand where they fail and how to improve recommendation quality for all items.

Concretely, much of the recent development in recommender systems research has been based on matrix factorization (MF) [17, 19, 32], where we use a database of user ratings of items to learn a latent bilinear model for predicting unobserved ratings. Following the Netflix Prize, these models generally aim to improve Root Mean Squared Error (RMSE) over a random holdout of ratings. While there are many advantages to such an approach, focusing on the average accuracy metrics, such as RMSE, leaves many items ill-served. In fact, as we will demonstrate, common properties of real-world data sets encourage a skewed recommendation policy where some items are modeled far worse than others.

Given this issue with classic factorization models, how can we learn a model, just using ratings data, that is focused on improving recommendation accuracy for badly-modeled items, or any subset of items? How can we use all observed ratings to improve the predictions of a subset? We call this the *focused learning* problem.

This problem is related to multiple previous lines of recommendation research. Research on the cold-start problem attempts to improve recommendation accuracy for items with few observed ratings, but often relies on side information, e.g., context [34] and review text [11], or probes users for more data [3, 2]. In our approach, we make use of *given rating data only*. Also related is research in transfer learn-

