Beyond Globally Optimal: Focused Learning for Improved Recommendations

Alex Beutel[1]*, Ed H. Chi[1], Zhiyuan Cheng[2]†, Hubert Pham[1], John Anderson[1]
[1]Google, Inc. [2]Pinterest
{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

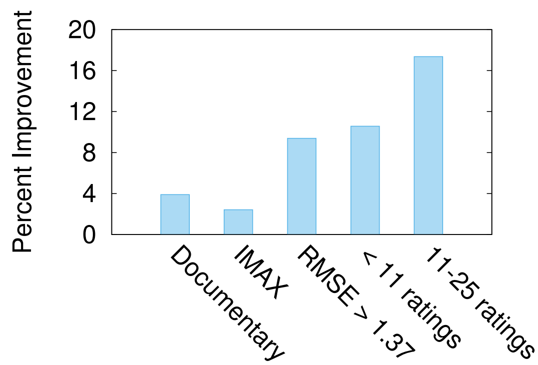

Figure 1: Focused learning improves MovieLens predictions for under-served categories of movies, including (1) niche genres, (2) items for which we have few observations, and (3) even the most badly-modeled items from our original model.

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 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-

*A portion of this work was done while at Carnegie Mellon University.
†A portion of this work was done while at Google, Inc.

WWW 2017, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4913-0/17/04.
http://dx.doi.org/10.1145/3038912.3052713

9. REFERENCES

[1] L. A. Adamic and B. A. Huberman. Power-law distribution of the world wide web. *Science*, 287(5461):2115–2115, 2000.

[2] M. Aharon, O. Anava, N. Avigdor-Elgrabli, D. Drachsler-Cohen, S. Golan, and O. Somekh. ExcUseMe: Asking users to help in item cold-start recommendations. In *RecSys*. ACM, 2015.

[3] O. Anava, S. Golan, N. Golbandi, Z. Karnin, R. Lempel, O. Rokhlenko, and O. Somekh. Budget-constrained item cold-start handling in collaborative filtering recommenders via optimal design. In *WWW*, pages 45–54, 2015.

[4] I. Barjasteh, R. Forsati, F. Masrour, A.-H. Esfahanian, and H. Radha. Cold-start item and user recommendation with decoupled completion and transduction. In *RecSys*, 2015.

[5] J. S. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl. Algorithms for hyper-parameter optimization. In *NIPS*, pages 2546–2554, 2011.

[6] A. Beutel, A. Ahmed, and A. J. Smola. ACCAMS: Additive Co-Clustering to Approximate Matrices Succinctly. In *WWW*, 2015.

[7] A. Beutel, K. Murray, C. Faloutsos, and A. J. Smola. CoBaFi: Collaborative Bayesian Filtering. In *WWW*, pages 97–108, 2014.

[8] L. Breiman and J. H. Friedman. Estimating optimal transformations for multiple regression and correlation. *JASA*, 80(391):580–598, 1985.

[9] E. Brochu, V. Cora, and N. De Freitas. A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv:1012.2599*, 2010.

[10] E. Christakopoulou and G. Karypis. Local item-item models for top-n recommendation. In *RecSys*, 2016.

[11] Q. Diao, M. Qiu, C.-Y. Wu, A. J. Smola, J. Jiang, and C. Wang. Jointly modeling aspects, ratings and sentiments for movie recommendation (JMARS). In *KDD*, pages 193–202. ACM, 2014.

[12] M. Dredze, K. Crammer, and F. Pereira. Confidence weighted linear classification. In *ICML*. ACM, 2008.

[13] A. M. Elkahky, Y. Song, and X. He. A multi-view deep learning approach for cross domain user modeling in recommendation systems. In *WWW*. ACM, 2015.

[14] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *SIGCOMM*, pages 251–262, 1999.

[15] P. Hamel. To have a tiger by the tail: Improving music recommendation for international users. In *ICML workshop*, 2015.

[16] L. Hu, J. Cao, G. Xu, L. Cao, Z. Gu, and C. Zhu. Personalized recommendation via cross-domain triadic factorization. In *WWW*. ACM, 2013.

[17] Y. Hu, Y. Koren, and C. Volinsky. Collaborative filtering for implicit feedback datasets. In *ICDM*, 2008.

[18] Y. Koren. Factorization meets the neighborhood: A multifaceted collaborative filtering model. In *KDD*, pages 426–434, New York, NY, USA, 2008.

[19] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *IEEE Computer*, 42(8):30–37, 2009.

[20] X. N. Lam, T. Vu, T. D. Le, and A. D. Duong. Addressing cold-start problem in recommendation systems. In *IMCOM*, pages 208–211. ACM, 2008.

[21] J. Lee, S. Kim, G. Lebanon, and Y. Singer. Local low-rank matrix approximation. In *ICML*, 2013.

[22] J. Lee, M. Sun, and G. Lebanon. PREA: Personalized recommendation algorithms toolkit. *JMLR*, 13(Sep), 2012.

[23] H. Ma, H. Yang, M. R. Lyu, and I. King. SoRec: social recommendation using probabilistic matrix factorization. In *CIKM*, pages 931–940. ACM, 2008.

[24] L. Mason, J. Baxter, P. L. Bartlett, and M. R. Frean. Boosting algorithms as gradient descent. *NIPS*, 2000.

[25] J. McAuley and J. Leskovec. Hidden factors and hidden topics: understanding rating dimensions with review text. In *RecSys*, pages 165–172. ACM, 2013.

[26] M. Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet mathematics*, 1(2):226–251, 2004.

[27] Y. J. Park and A. Tuzhilin. The long tail of recommender systems and how to leverage it. In *RecSys*, pages 11–18. ACM, 2008.

[28] S. Rendle. Factorization machines with libFM. *ACM TIST*, 3(3):57:1–57:22, May 2012.

[29] J. Riedl and J. Konstan. Movielens dataset, 1998.

[30] Y. Rong, X. Wen, and H. Cheng. A monte carlo algorithm for cold start recommendation. In *WWW*, pages 327–336. ACM, 2014.

[31] S. Sahebi and P. Brusilovsky. It takes two to tango: An exploration of domain pairs for cross-domain collaborative filtering. In *RecSys*. ACM, 2015.

[32] R. Salakhutdinov and A. Mnih. Bayesian probabilistic matrix factorization using markov chain monte carlo. In *ICML*, pages 880–887. ACM, 2008.

[33] M. Saveski and A. Mantrach. Item cold-start recommendations: learning local collective embeddings. In *RecSys*. ACM, 2014.

[34] A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock. Methods and metrics for cold-start recommendations. In *SIGIR*. ACM, 2002.

[35] L. Shi. Trading-off among accuracy, similarity, diversity, and long-tail: a graph-based recommendation approach. In *RecSys*. ACM, 2013.

[36] J. Snoek, H. Larochelle, and R. P. Adams. Practical bayesian optimization of machine learning algorithms. In *NIPS*, pages 2951–2959, 2012.

[37] C. Tan, E. H. Chi, D. Huffaker, G. Kossinets, and A. J. Smola. Instant foodie: Predicting expert ratings from grassroots. In *CIKM*, 2013.

[38] L. Vilnis and A. McCallum. Word representations via gaussian embedding. *arXiv:1412.6623*, 2014.

[39] H. Yin, B. Cui, J. Li, J. Yao, and C. Chen. Challenging the long tail recommendation. *Proc. VLDB Endow.*, 5(9):896–907, May 2012.

[40] X. Yu, H. Ma, B.-J. P. Hsu, and J. Han. On building entity recommender systems using user click log and freebase knowledge. In *WSDM*. ACM, 2014.

[41] Z. Zhao, Z. Cheng, L. Hong, and E. H. Chi. Improving user topic interest profiles by behavior factorization. In *WWW*, pages 1406–1416, 2015.