

Modularizing Deep Neural Network-Inspired Recommendation Algorithms

Longqi Yang
Cornell Tech, Cornell University
New York, NY, 10044
ly283@cornell.edu

Eugene Bagdasaryan
Cornell Tech, Cornell University
New York, NY, 10044
eb693@cornell.edu

Hongyi Wen
Cornell Tech, Cornell University
New York, NY, 10044
hw557@cornell.edu

ABSTRACT

This tutorial reviews recent developments of deep neural network-based recommendation algorithms and demonstrates how to extend and adapt such algorithms for diverse application scenarios. The customization is supported by OpenRec framework that modularizes neural recommenders. The tutorial consists of a lecture and two hands-on sessions. It targets intermediate and advanced audiences who already possess knowledge of deep neural networks and are interested in applying those knowledge to the domain of recommendation. Materials are available at: <http://openrec.ai/>

CCS CONCEPTS

• **Information systems** → **Collaborative filtering; Personalization; Recommender systems;**

KEYWORDS

Recommendation; Modularization; Deep Learning; OpenRec

ACM Reference Format:

Longqi Yang, Eugene Bagdasaryan, and Hongyi Wen. 2018. Modularizing Deep Neural Network-Inspired Recommendation Algorithms. In *Twelfth ACM Conference on Recommender Systems (RecSys '18)*, October 2–7, 2018, Vancouver, BC, Canada. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3240323.3241618>

1 INTRODUCTION

Deep Neural Network (DNN) [9]-inspired recommendation algorithms usually leverage numerous heterogeneous sub-models to incorporate a wide range of information, such as diverse user feedback signals and auxiliary, contextual, and cross-platform traces [2–5, 8, 13, 14, 17, 18]. These complexities pose significant challenges to the traditional recommender development practice which treats each algorithm as monolithic. As a result, in order to experiment with a new method for even a small part of an algorithm, or customize an algorithm for other scenarios, researchers and engineers need to re-implement the whole model from scratch or extensively patch existing code. For practitioners, this introduces significant barriers for applying state-of-the-art solutions to existing services.

To tackle these challenges, we present this tutorial to introduce and share hands-on experience with OpenRec [16], an open-source

tool that modularizes recommendation algorithms, such that a change, addition or deletion of a component does not affect others, and the development and testing can be readily achieved via plug-in and go. OpenRec aspires to democratize neural network-based recommendation solutions for people beyond experts.

2 TUTORIAL OUTLINE

This tutorial consists of a lecture (40 min) and two hands-on sessions (25 min each):

Lecture. The lecture (1) reviews matrix factorization-based and neural network-inspired recommendation algorithms, (2) motivates modularization of recommenders based on neural networks, and (3) introduces the architecture of OpenRec framework, as well as its key Python interfaces built on Tensorflow [1]. Specifically, the lecture covers following topics:

- *Classical matrix factorization based models*, including Bayesian Personalized Ranking (BPR) [12], Probabilistic Matrix Factorization (PMF) [11], and Weighted Approximately Ranked Pairwise loss (WARP) [15]. These methods lay foundations for the design of neural network-based recommenders.
- *Neural network-inspired models*: (1) models that analyze user-item interactions, including Collaborative Metric Learning (CML) [7], Neural Collaborative Filtering (NCF) [6], and Variational AutoEncoder [10], (2) models that incorporate auxiliary information, including Collaborative Deep Learning (CDL) [13], VBPR [5], and Latent Cross [2], and (3) end-to-end solutions, including wide and deep learning [3] and Youtube recommendations [4].
- *Motivations for modularization*. Many components of complex recommendation algorithms are reusable in different scenarios. Modularization can significantly reduce the development overhead and foster research on extensible and adaptable models.
- *OpenRec framework architecture*, including recommender, module (extraction, fusion, and interaction), and utility.
- *Tensorflow and OpenRec Python interfaces*. We briefly introduce the concept of computational graph used in Tensorflow, and additionally demonstrate OpenRec Python interfaces for model construction, training, and evaluation.

Hands-on sessions. We present two hands-on sessions to demonstrate how to use OpenRec to customize existing solutions, develop new algorithms, and conduct quick experiments. Jupyter notebooks are used for live programming. Detailed session information is presented below.

- *Experimenting existing neural recommenders*. In the first session, we demonstrate procedures for (1) training and testing

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

RecSys '18, October 2–7, 2018, Vancouver, BC, Canada

© 2018 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-5901-6/18/10.

<https://doi.org/10.1145/3240323.3241618>

existing neural network-based recommendation models, and (2) model and parameter selection.

- *Developing new algorithms through extension and adaptation.* In the second session, we show examples of developing (1) fairness-aware recommendation models, and (2) end-to-end solutions that explore heterogeneous information.

3 TARGETED AUDIENCE

This tutorial targets intermediate and advanced audiences, and requires basic knowledge of deep neural networks (such as MLP, CNN and RNN). We focus on recommendation-specific developments and applications of DNN. Prior experience with Tensorflow is not required. Basics of Tensorflow are reviewed in the tutorial.

4 IMPORTANCE OF THE TOPIC

Developing neural network-based algorithms for recommendation has shown great promise in recent years. Therefore, it is of significant utility for the Recsys community to have a tool that makes the development process easier, extensible, and reproducible. Therefore, experts can better build on each other's work and others can quickly prototype and deploy recommender systems. This is also a timely tutorial for two major Recsys trends: (1) developing deep learning algorithms for recommendations, and (2) conduct recommendations beyond user-item interaction signals (e.g., incorporating auxiliary user information, content analysis and contextual modeling).

5 PRESENTERS

Longqi Yang is a Ph.D. candidate of Computer Science at Cornell Tech, Cornell University. He is a member of the Connected Experiences lab and the small data lab where he is advised by Prof. Deborah Estrin. His current research focuses are user modeling, recommendation systems, and recommendation for social good. His work has been published and presented in top academic conferences, such as WWW, WSDM, Recsys, and CIKM. He co-organizes the workshop on "Immersive Recommendation: Deep User and Content Modeling for Personalization" at the NYC Media Lab annual summit 2017, and the workshop on "Machine Learning and Data Mining for Podcasts" at KDD 2018.

Eugene Bagdasaryan is a second-year Ph.D. student of Computer Science at Cornell Tech, Cornell University. His research focuses on privacy, modularity, and security of machine learning applications. He is a member of the Small Data Lab and possesses broad industrial experience.

Hongyi Wen is a third-year Ph.D. student of Information Science at Cornell Tech, Cornell University. He is a member of the Small Data Lab advised by Prof. Deborah Estrin. His research interests include personal data privacy-utility, user-centered computing for recommendation systems, and mobile health systems. His work has been published and presented in RecSys and CHI.

6 RELATED EVENTS

- *Immersive Recommendation: Deep User and Content Modeling for Personalization*, NYC Media Lab summit, 2017.
- *Harnessing and Customizing State-of-the-art Recommendation Solutions with OpenRec*, Strata Data Conference, 2018.

ACKNOWLEDGMENTS

This tutorial was funded by the National Science Foundation (Award #1700832) and Oath (the Connected Experiences Laboratory at Cornell Tech). It was further supported by the small data lab at Cornell Tech, which receives funding from NSF, NIH, RWJF, UnitedHealth Group, Google, and Adobe.

REFERENCES

- [1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. 2016. Tensorflow: a system for large-scale machine learning. In *OSDI*, Vol. 16. 265–283.
- [2] Alex Beutel, Paul Covington, Sagar Jain, Can Xu, Jia Li, Vince Gatto, and Ed H Chi. 2018. Latent Cross: Making Use of Context in Recurrent Recommender Systems. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. ACM, 46–54.
- [3] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishu Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, et al. 2016. Wide & deep learning for recommender systems. In *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*. ACM, 7–10.
- [4] James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, Blake Livingston, et al. 2010. The YouTube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems*. ACM, 293–296.
- [5] Ruining He and Julian McAuley. 2016. VBPR: Visual Bayesian Personalized Ranking from Implicit Feedback. In *AAAI* 144–150.
- [6] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 173–182.
- [7] Cheng-Kang Hsieh, Longqi Yang, Yin Cui, Tsung-Yi Lin, Serge Belongie, and Deborah Estrin. 2017. Collaborative metric learning. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 193–201.
- [8] Cheng-Kang Hsieh, Longqi Yang, Honghao Wei, Mor Naaman, and Deborah Estrin. 2016. Immersive recommendation: News and event recommendations using personal digital traces. In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 51–62.
- [9] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature* 521, 7553 (2015), 436.
- [10] Dawen Liang, Rahul G Krishnan, Matthew D Hoffman, and Tony Jebara. 2018. Variational Autoencoders for Collaborative Filtering. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 689–698.
- [11] Andriy Mnih and Ruslan R Salakhutdinov. 2008. Probabilistic matrix factorization. In *Advances in neural information processing systems*. 1257–1264.
- [12] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*. AUAI Press, 452–461.
- [13] Hao Wang, Naiyan Wang, and Dit-Yan Yeung. 2015. Collaborative deep learning for recommender systems. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1235–1244.
- [14] Honghao Wei, Cheng-Kang Hsieh, Longqi Yang, and Deborah Estrin. 2016. GroupLink: Group Event Recommendations Using Personal Digital Traces. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*. ACM, 110–113.
- [15] Jason Weston, Samy Bengio, and Nicolas Usunier. 2011. Wsabie: Scaling up to large vocabulary image annotation. In *IJCAI*, Vol. 11. 2764–2770.
- [16] Longqi Yang, Eugene Bagdasaryan, Joshua Gruenstein, Cheng-Kang Hsieh, and Deborah Estrin. 2018. OpenRec: A Modular Framework for Extensible and Adaptable Recommendation Algorithms. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining (WSDM '18)*. ACM, New York, NY, USA, 664–672. <https://doi.org/10.1145/3159652.3159681>
- [17] Longqi Yang, Chen Fang, Hailin Jin, Matthew D Hoffman, and Deborah Estrin. 2017. Personalizing Software and Web Services by Integrating Unstructured Application Usage Traces. In *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 485–493.
- [18] Longqi Yang, Cheng-Kang Hsieh, and Deborah Estrin. 2015. Beyond classification: Latent user interests profiling from visual contents analysis. In *Data Mining Workshop (ICDMW), 2015 IEEE International Conference on*. IEEE, 1410–1416.