Deep Learning for Search and Recommender Systems in Practice

Zhoutong Fu, Huiji Gao, Weiwei Guo, Sandeep Kumar Jha, Jun Jia*, Xiaowei Liu, Bo Long, Jun Shi, Sida Wang, Mingzhou Zhou
LinkedIn Corporation, Mountain View, California
{zfu,hgao,wguo,sjha,jjia*,xwli,blong,jshi,sidwang,mizhou}@linkedin.com

ABSTRACT

In this talk, we will go over the components of personalized search and recommender systems and demonstrate the applications of various deep learning techniques along the way.

Search and recommender systems are probably the most prevalent ML powered application across the industry. They share most of the components composition and provide a user a ranked list of items, while there is subtle difference that a search system typically acts passively with a clear user intention in terms of queries and a recommender system acts more proactively.

Deep learning has been wildly successful in solving complex tasks such as image recognition, speech recognition, natural language processing and understanding, machine translation, etc. In the area of personalized recommender systems, deep learning has been showing tremendous impact in recent years.

Search and recommender systems can be staged roughly in three phases: 1. User and query understanding, where a query or a user profile are processed so that the systems can use the processed information to 2. retrieve all the related items (high recall) and 3. rank the items by the order of the most relevance to the user's intent (high precision). Each phase has its unique challenges but deep learning has been ubiquitously pushing beyond the limit.

After walking through the talk, we hope the audience would gain some first-hand experience building a personalized search/recommender system using deep learning techniques.

KEYWORDS

Deep Learning, Search Engine, Recommender System, Natural Language Process, Ranking System

1 OUTLINE

The presentation will be divided into four sessions and last 6 hours in total.

1.1 Introduction to deep learning, search and recommender systems

- System architectural overview
- Major components of search and recommender systems and common approaches
- Deep learning and and its applications in search and recommender systems

1.2 Understanding

(1) User and query understanding
- Query understanding infers the intent of a search engine user by extracting semantic meaning from the searcher's keywords.
- User understanding provides personalization features for candidate retrieval and ranking.

(2) Hands-on session
- Train a query intent model.
- Train a query auto completion model.

1.3 Candidate Retrieval

(1) Candidate retrieval for a search system through indices
- For a search system (search engine), the candidate selection is typically handled by a reverted index.

(2) Candidate selection for a recommender system
- A general recommender system usually has multiple sources of candidates.

(3) Hands-on session:
- Setup ElasticSearch with pre-populated indices;
- Train a deep KNN model that can be used for candidate selection of a recommender system.

1.4 Ranker and Re-ranker

(1) Learning to Rank (LTR) in search and recommender systems
- Different strategies applied to a ranking problem, including point-, pair- and list-wise ranking algorithms.
- Business rule-based mixer as a special kind of reranker.

(2) Hands-on session:
- Train a Generalized Deep Mixed Model(GDMix, an extension to GLMix [22] to be open-sourced by LinkedIn) with DeText [14] and serve with ElasticSearch;
- Train and serve a random effect TensorFlow.js model as a re-ranker in a browser.

2 PRESENTERS' BIOGRAPHY

Dr. Huiji Gao leads the AI Algorithms Foundation team at LinkedIn. He has broad interests in machine learning/AI and its applications, including search/recommender systems, computational advertising, and NLP. He received Ph.D. in Computer Science from Arizona State University, and B.S./M.S. from Beijing University of Posts and Telecommunications. He has filed over 10 U.S. patents and published 40 publications in top journals and conferences including KDD, AAAI, WWW, ICDM, DMKD with thousands of citations.

*Corresponding author.

KDD '20, August 23–27, 2020, Virtual Event, CA, USA

ACM ISBN 978-1-4503-7998-4/20/08.
https://doi.org/10.1145/3394486.3406709

Dr. Jun Jia is a Senior Staff Software Engineer in the Artificial Intelligence team at LinkedIn where he works on developing state-of-the-art algorithms to improve LinkedIn's Search and Recommendation systems. He has published many peer-reviewed papers and served as reviewers in journals and conferences. Prior to LinkedIn, he obtained his MS and PhD in CS and Math from the University of North Carolina at Chapel Hill and worked as a technical staff member at ORNL.

Dr. Jun Shi is a staff software engineer at LinkedIn, where he leads various efforts on promoting natural language processing in search with deep learning technologies. His research interest lies in the area of machine learning with emphasis on natural language processing. He was an author of CaffeOnSpark and TensorflowOnSpark. He was a contributor to Tensorflow and created verbs interface for Tensorflow. Jun Shi received a doctoral degree in Electrical Engineering from UCLA. He was a co-recipient of 2009 IEEE Communications Society & Information Theory Society Joint Paper Award.

Dr. Xiaowei Liu is a senior software engineer in the NLP team at LinkedIn where she focuses on applying state-of-the-art NLP algorithms to power and improve LinkedIn products. She received her Ph.D. in EE from Stony Brook University, and B.S. from Beijing University of Posts and Telecommunications.

Dr. Bo Long leads LinkedIn's AI Foundations team. He also worked at Particle Media, Yahoo! Labs, IBM Watson and Google Lab. He has 15 years of experience in data mining and machine learning with applications to web search, recommendation, and social network analysis. He holds dozens of innovations and has published peer-reviewed papers in top conferences and journals including ICML, KDD, ICDM, AAAI, SDM, CIKM, and KAIS. He has served as reviewers, workshops co-organizers, conference organizer committee members, and area chairs for multiple conferences, including KDD, NIPS, SIGIR, ICML, SDM, CIKM, JSM etc

Dr. Mingzhou Zhou is a senior software engineer at LinkedIn. His work focuses on incorporating deep neural network model to improve LinkedIn's large scale recommendation and search system.

Dr. Weiwei Guo leads the NLP team at LinkedIn. He obtained his Ph.D. with a focus on NLP from Columbia University, and B.S. from Sun Yat-sen University. Weiwei has published over 20 peer-reviewed papers in top conferences including ACL, EMNLP, NAACL, SIGIR, KDD with 1000+ citations.

Sandeep Jha is a Staff Technical Program Manager in the Artificial Intelligence group at LinkedIn, where he leads programs that empowers LinkedIn's Search and Recommendation systems. Before LinkedIn, he was a Sr. Technical Program Manager at Amazon, where he worked on improving the search result in the first page of Amazon worldwide. Before that, he worked at Facebook, where he led initiatives to enhance ad quality and launch of Facebook Marketplace and Instagram Shopping.

REFERENCES

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mane, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viegas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2016. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. arXiv:cs.DC/1603.04467

[2] Deepak Agarwal, Bee-Chung Chen, Rupesh Gupta, Joshua Hartman, Qi He, Anand Iyer, Sumanth Kolar, Yiming Ma, Pannagadatta Shivaswamy, Ajit Singh, and Liang Zhang. 2014. Activity ranking in LinkedIn feed. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (08 2014). https://doi.org/10.1145/2623330.2623362

[3] Deepak Agarwal, Liang Zhang, Bee-Chung Chen, Qi He, Zhenhao Hua, Guy Lebanon, Yiming Ma, Pannagadatta Shivaswamy, Hsiao-Ping Tseng, and Jaewon Yang. 2015. Personalizing LinkedIn Feed. 1651–1660. https://doi.org/10.1145/2783258.2788614

[4] Trapit Bansal, David Belanger, and Andrew McCallum. 2016. Ask the gru: Multi-task learning for deep text recommendations. In *RecSys*.

[5] Leonid Boytsov, David Novak, Yury Malkov, and Eric Nyberg. 2016. Off the Beaten Path: Let's Replace Term-Based Retrieval with k-NN Search. In *CIKM*.

[6] Christopher Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Gregory Hullender. 2005. Learning to Rank using Gradient Descent. *ICML 2005 - Proceedings of the 22nd International Conference on Machine Learning*, 89–96. https://doi.org/10.1145/1102351.1102363

[7] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. 785–794. https://doi.org/10.1145/2939672.2939785

[8] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, G.s Corrado, Wei Chai, Mustafa Ispir, Rohan Anil, Zakaria Haque, Lichan Hong, Vihan Jain, Xiaobing Liu, and Hemal Shah. 2016. Wide & Deep Learning for Recommender Systems. 7–10. https://doi.org/10.1145/2988450.2988454

[9] Corinna Cortes and Vladimir Vapnik. 1995. Support Vector Network. *Machine Learning* 20 (09 1995), 273–297. https://doi.org/10.1007/BF00994018

[10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL*.

[11] Nadia Fawaz, Saurabh Kataria, Benjamin Le, Liang Zhang, and Ganesh Venkataraman. 2017. Deep Learning for Personalized Search and Recommender Systems. In *KDD*.

[12] Homa B. Hashemi, Amir Asiaee, and Reiner Kraft. 2016. Query Intent Detection using Convolutional Neural Networks. In *International Conference on Web Search and Data Mining, Workshop on Query Understanding*.

[13] Baotian Hu, Zhengdong Lu, Hang Li, , and Qingcai Chen. 2014. Convolutional Neural Network Architectures for Matching Natural Language Sentences. In *NIPS*.

[14] LinkedIn. 2020. Deep neural ranking framework with Text understanding. https://github.com/linkedin/detext

[15] Bhaskar Mitra, Fernando Diaz, , and Nick Craswell. 2017. Learning to match using local and distributed representations of text for web search. In *WWW*.

[16] Yelong Shen, Xiaodong He, Jianfeng Gao, Li Deng, and Grégoire Mesnil. 2014. Learning deep structured semantic models for web search using clickthrough data. In *WWW*.

[17] Yangyang Shi, Kaisheng Yao, Le Tian, and Daxin Jiang. 2016. Deep LSTM based feature mapping for query classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

[18] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *NIPS*.

[19] Jun Xu, Xiangnan He, and Hang Li. 2018. Deep Learning for Matching in Search and Recommendation. In *SIGIR*.

[20] Hamed Zamani, Mostafa Dehghani, W. Bruce Croft, Erik Learned-Miller, and Jaap Kamps. 2018. From Neural Re-Ranking to Neural Ranking: Learning a Sparse Representation for Inverted Indexing. In *CIKM*.

[21] Hamed Zamani, Bhaskar Mitra, Xia Song, Nick Craswell, and Saurabh Tiwary. 2018. Neural ranking models with multiple document fields. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*.

[22] XianXing Zhang, Yitong Zhou, Yiming Ma, Bee-Chung Chen, Liang Zhang, and Deepak Agarwal. 2016. GLMix: Generalized Linear Mixed Models For Large-Scale Response Prediction. 363–372. https://doi.org/10.1145/2939672.2939684