Deep Transfer Learning for Search and Recommendation

Yanen Li, Yang Yang, Sen Zhou, Jian Qiao, Bo Long
LinkedIn, Mountain View, California
{yali,yyang,sezhou,jqiao,blong}@linkedin.com

ABSTRACT

Training data sparsity is a common problem for many real-world applications in Search and Recommendation domains. Even for applications with a lot of training data, in the cold-start scenario we usually do not get enough labeled data. Transfer Learning is a promising approach for addressing this problem. In addition, features might interact with each other in a complex way that traditional approaches might not be able to represent, Deep Transfer Learning, which leverages Deep Neural Networks for Transfer Learning, might be able to catch such deep patterns hidden in complex feature interactions. Due to these reasons, recently Deep Transfer Learning research has gained a lot of attention and has been successfully applied to many real-world applications. This tutorial offers an overview of Deep Transfer Learning approaches in Search and Recommendation domains from the industry perspective. In this tutorial We first introduce the basic concepts and major categories of Deep Transfer Learning. Then we focus on recent developments of Deep Transfer Learning approaches in the Search and Recommendation domains. After that we will introduce two real-world examples of how to apply Deep Transfer Learning methods to improve Search and Recommendation performance at LinkedIn. Finally we will conclude the tutorial with discussion of future directions.

KEYWORDS

Deep Transfer Learning, Search and Recommendation, Unified Embeddings

ACM Reference Format:
Yanen Li, Yang Yang, Sen Zhou, Jian Qiao, Bo Long. 2020. Deep Transfer Learning for Search and Recommendation. In *Companion Proceedings of the Web Conference 2020 (WWW '20 Companion), April 20–24, 2020, Taipei, Taiwan.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3366424.3383115

1 INTRODUCTION

Search and recommendation applications usually suffer from training data sparsity problems, especially for small to mid-size verticals where user feedbacks are hard to obtained. Even for applications with a lot of training data, in the cold-start scenario we usually do not get enough labeled data. Transfer Learning has been demonstrated as an effective way for addressing this problem [9]. In addition, from the perspective of training more effective models, features might interact with each other in a complex way that traditional approaches might not be able to represent. Deep Transfer Learning might be able to capture such deep features interactions that cannot be obtained by traditional methods. For these reasons, recently Deep Transfer Learning research has gained a lot of attention both in academia and industry.

In this tutorial we summarize the recent advancements in Deep Transfer Learning for search and recommendation domains. We first give an overview of the Search and Recommendation and Transfer Learning approaches [9, 17] in these domains. Then we introduce the basic concepts of Deep Transfer Learning [14, 20], including major types of Deep Transfer Learning, major methodologies and their applications to Search and Recommendation applications. After that, we share our hands-on experience by leveraging Deep Transfer Learning approaches for LinkedIn use-cases. In the end, we conclude our tutorial with some important future trends.

Our tutorial will focus on the following topics of Deep Transfer Learning in search and recommendation: 1) **Domain Adaptation** methods attempt to learn a well performing model from source distribution to the target distribution. Recent domain adaptation methods learn deep neural transformations that map both domains into a common feature space, or tries to select high-quality training data from source domain that can augment the data in target domain [1, 2, 18, 21]; 2) **Multi-task Deep Learning** [11, 12] tries to learn unified entity [22] or text embeddings [24] via combining cross-domain training data with shared network structure and parameters [6] and task specific user behavior such as sequential user feed-backs [5, 8, 10]; 3) **Model Distillation** is an effective model compression method in which a small model is trained to mimic a pre-trained, larger model [3, 4, 6, 15, 16, 23]; 4) **Privacy-preserving transfer learning** e.g. Federated Transfer Learning builds effective models while preserving most of the privacy of the data [7, 13, 19]. At last, we illustrate end-to-end real-world examples leveraging Deep Transfer Deep approaches in LinkedIn, sharing our experiences in algorithm development and infrastructure building, especially on a) Learning Unified Embeddings by Deep Transfer Learning and b) Adversary Network-based Selective Deep Transfer Learning for Data Augmentation.

Targeted Audience: The tutorial targets at a broad audience from Search and Recommendation and other related areas, including academic and industrial researchers, graduate students, and practitioners. After the tutorial, we expect the audience could grasp basic concepts and guidelines of applying state-of-the-art Deep Transfer Learning approaches in search recommendation applications, and gained real-world experiences through illustrative examples of end-to-end examples.

Interaction style: This tutorial is mostly conducted by presentations.

WWW '20 Companion, April 20–24, 2020, Taipei, Taiwan

ACM ISBN 978-1-4503-7024-0/20/04.
https://doi.org/10.1145/3366424.3383115

2 TUTORIAL OUTLINE (3.5 HOURS IN TOTAL)

2.1 Introduction (30 mins)

(1) Overview of Search and Recommendation
(2) Transfer Learning in Search and Recommendation

2.2 Deep Transfer Learning (30 mins)

(1) Preliminaries
(2) Problem Settings of Deep Transfer Learning
- Inductive Transfer Learning
- Transductive Transfer Learning
- Unsupervised Transfer Learning

(3) Common Approaches of Deep Transfer Learning
- Instance based Deep Transfer Learning
- Pre-training and Model Fine-tuning
- Multi-task Learning
- Model Distillation
- Privacy-preserving Transfer Learning

2.3 Deep Transfer Learning in Search and Recommendation (90 mins)

(1) Domain Adaptation
- Adversarial Discriminative Domain Adaptation
- Instanced-based Selective Deep Transfer Learning using Adversarial Networks

(2) Learning Unified Embeddings by Multi-task Learning
- Multi-task Learning for Unified Entity and Text Embeddings
- Multi-task Learning for multi-type user behavior

(3) Model Distillation
- Modeling Compression by Model Distillation for Search and Recommendation

(4) Privacy-preserving transfer learning

3 CASE-STUDY (50 MINS)

(1) An End-to-end Example of Learning Unified Embeddings by Deep Transfer Learning at LinkedIn
(2) Adversary Network-based Selective Deep Transfer Learning for Data Augmentation for LinkedIn Applications

4 FUTURE TRENDS AND CONCLUSIONS (10 MINS)

REFERENCES

[1] Zhangjie Cao, Mingsheng Long, Jianmin Wang, and Michael I. Jordan. 2018. Partial Transfer Learning With Selective Adversarial Networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[2] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor S. Lempitsky. 2015. Domain-Adversarial Training of Neural Networks. *J. Mach. Learn. Res.* 17 (2015), 59:1–59:35.

[3] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. 2015. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531* (2015).

[4] Yoon Kim and Alexander M Rush. 2016. Sequence-level knowledge distillation. *arXiv preprint arXiv:1606.07947* (2016).

[5] Jing Li, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Tao Lian, and Jun Ma. 2017. Neural Attentive Session-based Recommendation. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM '17)*. ACM, New York, NY, USA, 1419–1428. https://doi.org/10.1145/3132847.3132926

[6] Xiaodong Liu, Pengcheng He, Weizhu Chen, and Jianfeng Gao. 2019. Improving Multi-Task Deep Neural Networks via Knowledge Distillation for Natural Language Understanding. *arXiv preprint arXiv:1904.09482* (2019).

[7] Yang Liu, Tianjian Chen, and Qiang Yang. 2018. Secure Federated Transfer Learning. *CoRR* abs/1812.03337 (2018). arXiv:1812.03337 http://arxiv.org/abs/1812.03337

[8] Xiao Ma, Liqin Zhao, Guan Huang, Zhi Wang, Zelin Hu, Xiaoqiang Zhu, and Kun Gai. 2018. Entire space multi-task model: An effective approach for estimating post-click conversion rate. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, 1137–1140.

[9] S. J. Pan and Q. Yang. 2010. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering* 22, 10 (Oct 2010), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

[10] Kan Ren, Jiarui Qin, Yuchen Fang, Weinan Zhang, Lei Zheng, Weijie Bian, Guorui Zhou, Jian Xu, Yong Yu, Xiaoqiang Zhu, and Kun Gai. 2019. Lifelong Sequential Modeling with Personalized Memorization for User Response Prediction. In *SIGIR*.

[11] Sebastian Ruder. 2017. An Overview of Multi-Task Learning in Deep Neural Networks. *CoRR* abs/1706.05098 (2017). arXiv:1706.05098 http://arxiv.org/abs/1706.05098

[12] Ozan Sener and Vladlen Koltun. 2018. Multi-Task Learning as Multi-Objective Optimization. In *Advances in Neural Information Processing Systems 31*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Eds.). Curran Associates, Inc., 527–538. http://papers.nips.cc/paper/7334-multi-task-learning-as-multi-objective-optimization.pdf

[13] Virginia Smith, Chao-Kai Chiang, Maziar Sanjabi, and Ameet S Talwalkar. 2017. Federated Multi-Task Learning. In *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). Curran Associates, Inc., 4424–4434. http://papers.nips.cc/paper/7029-federated-multi-task-learning.pdf

[14] Chuanqi Tan, Fuchun Sun, Tao Kong, Wenchang Zhang, Chao Yang, and Chunfang Liu. 2018. A Survey on Deep Transfer Learning. *ArXiv* abs/1808.01974 (2018).

[15] Yong Kiam Tan, Xinxing Xu, and Yong Liu. 2016. Improved recurrent neural networks for session-based recommendations. In *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*. ACM, 17–22.

[16] Jiaxi Tang and Ke Wang. 2018. Ranking distillation: Learning compact ranking models with high performance for recommender system. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2289–2298.

[17] Lisa Torrey and Jude Shavlik. 2010. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. IGI Global, 242–264.

[18] Bo Wang, Minghui Qiu, Xisen Wang, Yaliang Li, Yu Gong, Xiaoyi Zeng, Jun Huang, Bo Zheng, Deng Cai, and Jingren Zhou. 2019. A Minimax Game for Instance Based Selective Transfer Learning. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19)*. ACM, New York, NY, USA, 34–43. https://doi.org/10.1145/3292500.3330841

[19] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. 2019. Federated Machine Learning: Concept and Applications. *ACM Trans. Intell. Syst. Technol.* 10, 2, Article 12 (Jan. 2019), 19 pages. https://doi.org/10.1145/3298981

[20] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. 2014. How transferable are features in deep neural networks?. In *NIPS*.

[21] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. 2014. How Transferable Are Features in Deep Neural Networks?. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'14)*. MIT Press, Cambridge, MA, USA, 3320–3328. http://dl.acm.org/citation.cfm?id=2969033.2969197

[22] Andrew Zhai, Hao-Yu Wu, Eric Tzeng, Dong Huk Park, and Charles J. Rosenberg. 2019. Learning a Unified Embedding for Visual Search at Pinterest. *ArXiv* abs/1908.01707 (2019).

[23] Ying Zhang, Tao Xiang, Timothy M. Hospedales, and Huchuan Lu. 2018. Deep Mutual Learning. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[24] Jinfeng Zhuang and Yu Liu. 2019. PinText: A Multitask Text Embedding System in Pinterest. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19)*. ACM, New York, NY, USA, 2653–2661. https://doi.org/10.1145/3292500.3330671