Graph Embeddings for Web-scale Recommender Systems

Ahmed El-Kishky
aelkishky@twitter.com
Twitte Cortex
Seattle, WA, USA

Michael Bronstein
mbronstein@twitter.com
University of Oxford and Twitter Cortex
Oxford, UK

Ying Xiao
yxiao@twitter.com
Twitter Cortex
San Francisco, CA, USA

Aria Haghighi
ahaghighi@twitter.com
Twitter Cortex
Seattle, WA, USA

ABSTRACT

Recommender systems are fundamental building blocks of modern consumer web applications that seek to predict user preferences to better serve relevant items. As such, high-quality user and item representations as inputs to recommender systems are crucial for personalized recommendation. To construct these user and item representations, self-supervised graph embedding has emerged as a principled approach to embed a variety of graphs such as user social graphs, user membership graphs, user-item engagements, and other heterogeneous graphs. In this tutorial we discuss different families of approaches to self-supervised graph embedding. Within each family, we outline a variety of techniques, their merits and disadvantages, and expound on latest works. Finally, we demonstrate how to effectively utilize the resultant large embedding tables to improve candidate generation and ranking in modern industry-scale deep-learning recommender systems.

1 INTRODUCTION

Motivation

Large pretrained embedding tables have become an important tool in recommender systems; they have allowed ML practitioners to understand both user behavior and item appeal. These pretrained representations have been used to improve recall in lightweight candidate generation and, when used as features, they have been shown to improve ranking models. As such, pretrained embeddings are widely used in a variety of industry recommender systems such as for app and video recommendation for Google Play and Youtube, personalized search at AirBnB, pin recommendation at Pinterest, connecting users based on interest at Etsy, ads recommendation at Twitter, and news article recommendation at Yahoo! Japan.

Despite their prevalence in recommender systems, the plethora of techniques used to learn pretrained embedding tables for recommender systems appear disparate or ad-hoc. As such, it may be difficult for recommender systems practitioners to determine how to best leverage abundant data for self-supervised pretraining and which self-supervised task is best to learn pretrained embeddings. Despite a scattered literature of pretraining approaches, many self-supervised embedding approaches can be viewed from the lens of *graph embedding*. Different graph embedding approaches provide the tools and formalization to utilize a wide-variety of available graph and network data to learn embeddings for use in recommender systems. A principled outline can demystify both how to train large-scale graph embeddings as well as how to utilize them for recommender systems tasks such as candidate generation and ranking.

Objectives

The objective of this tutorial is to introduce the audience to the task of large-scale graph embeddings with an application to improving end-to-end recommender systems. Tailoring our tutorial to audience members of different levels of expertise, we will introduce the subfield of recommender systems, and guide the audience to understanding the importance of pretrained representations to modern industry recommender systems. Following this, we will provide the audience to a collection of graph embedding techniques, their motivations, strengths and weaknesses. We will conclude with a comprehensive overview of how to integrate pretrained graph embeddings within recommender systems. We hope this tutorial spurs the machine learning and data mining community to develop scalable techniques to tackle the interesting challenges in large-scale graph pretraining and develop innovative ways to utilize them in web-scale recommender systems.

What will be covered?

Preliminaries: We introduce the audience to the broad subject of learning pretrained embeddings by providing motivation in the context of leveraging pretrained user and item embeddings for industry recommender systems. Within this context, we introduce graph embedding as a framework for leveraging various relational network data to learn rich dense representations. Finally, we describe at a high-level the end-to-end recommendation pipeline and describe how large embedding tables learned via graph embedding methods can improve end-to-end recommender systems – from unsupervised candidate generation to supervised ranking models.

Homogeneous Graphs:
In this part of the presentation, we first begin by discussing a variety of logged data such as social networks and engagement graphs and how they can be represented as graphs. Starting with these homogeneous graphs (i.e., a single node and edge type), we introduce several self-supervised techniques that learn a low-dimensional representation for nodes in a graph. More specifically we outline three families of homogenous embeddings (1) random walk variants with SkipGram (2) embeddings with high-order proximity and (3) deep neural network embeddings.

Heterogeneous Graphs: In this part of our presentation, we introduce the notion of heterogeneous graphs (a.k.a heterogeneous networks). We formulate heterogeneous graph embeddings and outline types of algorithms for embedding them. More specifically we discuss (1) enriching graph nodes with textual content (2) heterogeneous information network embeddings and (3) knowledge graph embeddings to embed multi-type, multi-relation networks.

Graph Neural Networks: In this part of our presentation, we discuss approaches to embedding via graph neural networks (GNN) – neural networks that capture topology and relationships within graph. We begin with an overview of GNNs and dive deeper into three approaches (1) applying message-passing algorithms between nodes of the graph (2) convolutional GNNs and (3) attention-based GNNs. Following this, we discuss scalable state-of-the art approaches, and approaches that have been applied in industry settings to recommender systems.

Pretrained Embeddings for Recommender Systems: In the final part of our presentation, we discuss recommender systems and how pretrained embedding tables can be integrated to improve recommendation. We start with how pretrained embeddings can be utilized in light-weight candidate generation strategies. We then outline popular deep-learning recommender systems and demonstrate the value of pretrained embeddings for them and how to effectively incorporate them. We touch on important aspects such as fine-tuning, scalability, and how to utilize large embeddings in latency-critical recommender systems.

2 WHY A TUTORIAL AT ECML-PKDD 2022

With the proliferation online marketplaces, social networks, and other information services, providing relevant content and items to users is of utmost importance. For these and other online services, machine-learning based recommender systems drive engagement by serving up relevant content to users. However, supervised machine-learning recommender systems can demonstrate poor performance when training data is sparse. In the extreme case, performance can greatly degrade when providing recommendations for new items or users due to lack of information. Without sufficient data, the model parameters cannot be well estimated and users' preference and item appeal cannot be well modeled.

It has been shown that the data sparsity problem in recommender systems can be alleviated by transferring knowledge from other domains or tasks. This is relevant to ECML-PKDD as machine learning and data mining is crucial in gathering, organizing, and extracting value in the form of pretrained representations for transfer learning. In this tutorial we introduce and outline recent and high-impact research for transfer learning via large-scale graph embedding pretraining and describe effective methodologies to leverage these embeddings to improve recommender systems.

Audience and Prerequisites

Researchers and practitioners in the field of data mining, machine learning, and more specifically recommender systems. While an audience with a good background in these areas would benefit most from this tutorial, we believe the material to be presented would give a general audience and newcomers an introductory pointer to the current work and important research topics in this subfield, and inspire them to learn more. Only preliminary knowledge about machine learning, data mining and their applications are needed.

3 TUTORIAL OUTLINE

This tutorial presents a comprehensive overview of the techniques developed for graph embeddings with a focus on leveraging graph embeddings for recommender systems. On this topic, we will discuss the following key issues and papers.

(1) **Motivation & Background** (15 minutes)
  (a) Recommender systems preliminary overview.
  (b) Sparsity and cold-start in recommender systems.
  (c) How pretrained embeddings can improve recommendation.
(2) **Homogeneous Graph Embeddings** (30 minutes)
  (a) Random-walk approaches
    (i) DeepWalk
    (ii) node2Vec
  (b) Incorporating high-order proximity
    (i) LINE embeddings
    (ii) SDE embeddings
    (iii) GraRep embeddings
  (c) Deep neural networks (stacked denoising autoencoders)
(3) **Heterogeneous Embeddings** (60 minutes)
  (a) Content-enriched node representations
  (b) Heterogeneous information network (HIN) Embeddings embeddings
  (c) Knowledge graph embeddings
    (i) Translating Embeddings (TransE)
    (ii) Complex Embeddings (ComplEx)
    (iii) Bilinear Embeddings (DistMult)
(4) **Break** (30 minutes)
(5) **Graph Neural Networks** (60 minutes)
  (a) GNN Overview
    (i) Convolutional GNNs
    (ii) Attention-based GNNs
    (iii) Message-passing GNNs
  (b) Scalable GNNs for recommendation
    (i) GraphSage
    (ii) PinSage
    (iii) SIGN: Scalable Inception Graph Neural Networks
    (iv) AliGraph
(6) **Pretrained Embeddings in Recommender Systems** (45 minutes)
  (a) Candidate Generation:
    (i) PinnerSage
    (ii) knn-Embed
    (iii) Deep Candidate Generation
  (b) Ranking Models
    (i) Wide and Deep model
    (ii) Deep and Cross model
    (iii) Deep Learning Recommendation Model (DLRM)
  (c) Embedding tables in ranking models
    (i) Deep Hash Embeddings (DHE) .

(ii) Task-specific fine-tuning embeddings
(iii) Mixtures of embeddings

Supplementary Material: In addition to the lecture-stle tutorial, we will provide easy-to-run Jupyter notebook examples of a select number of the graph-based embedding approaches as well as several recommender system implementation that leverage these embeddings.

4 INSTRUCTORS

The tutorial will be presented by Aria Haghighi, Ahmed El-Kishky, Michael Bronstein, and Ying Xiao.

- **Aria Haghighi** is a senior ML manager at Twitter Cortex where he supports content understanding efforts to improve Twitter products. Before that, Haghighi has worked on machine learning products at Facebook, Apple, and startups. He received his PhD from UC Berkeley, where he was supported by a Microsoft Research Fellowship and won multiple best paper awards at the North American Association of Computational Linguistics (NAACL) conference. Haghighi has published papers at venues such as ACL, EMNLP, SIGMOD, NAACL, ICML, Computational Linguistics, and others. He will be presenting the portions of the tutorial covering the motivation, background, and embedding homogeneous graphs.
- **Ahmed El-Kishky** is a researcher at Twitter Cortex where he works on developing scalable techniques for self-supervised representation learning and applying them to web-scale recommender systems. Before that, El-Kishky was a researcher at Facebook AI where he developed new techniques in multilingual natural language processing. El-Kishky received his PhD from the University of Illinois at Urbana-Champaign where he was supported by the National Science Foundation Graduate Research Fellowship (NSF-GRF) and the National Defense Science and Engineering Graduate (NDSEG) Fellowship. In his career, El-Kishky has published papers and given tutorials in venues such as JMLR, KDD, SIGIR, ACL, EMNLP, EACL SIGIR, WMT, VLDB, SIGMOD, WWW, WSDM, ICDM, and AACL. He will be presenting the portions of the tutorial covering embedding heterogeneous graphs.
- **Michael Bronstein** is the DeepMind Professor of AI at the University of Oxford and Head of Graph Learning Research at Twitter, where he works on efficient graph ML models and their applications. He previously held visiting appointments at Stanford, Harvard, MIT, TUM, and the Institute for Advanced Study in Princeton. He is a Member of the Academia Europaea, Fellow of IEEE, ELLIS, IAPR and BCS, ACM Distinguished Speaker, TEDx speaker, World Economic Forum Young Scientist and speaker, and the recipient of the Royal Society Wolfson Research Merit Award, five ERC grants, two Google Faculty Awards, two Amazon AWS ML Research Awards, and TED Audacious Prize (as part of Project CETI). Michael published papers in main ML venues such as NeurIPS, ICML, ICLR, JMLR, and PAMI, gave multiple talks on graph neural networks (including a keynote at ICLR 2021), and taught multiple tutorials and short courses on the topic (including at NeurIPS and several MLSS summer schools). He will be presenting the portions of the tutorial covering graph neural networks.
- **Ying Xiao** is a researcher at Twitter Cortex where he works on architectures and scaling up recommender systems. Before that, Xiao was a researcher at Google Research where he worked on Hessian eigenspectra, text-to-speech, and optical character recognition. Xiao received his PhD from the Georgia Institute of Technology. Throughout his career, he has published papers and organized workshops at ICML, ICLR, CVPR, ICASSP, STOC, and COLT. He will be presenting the portions of the tutorial covering how to leverage pretrained graph embeddings for recommender systems.

5 INSTRUCTORS' PREVIOUS TUTORIALS

The authors have presented many tutorials at machine learning, natural language processing, data mining and database conferences. The following is a set of recent tutorials given by the authors.

- **Tutorial at Nepal Winter School in AI, Kathmandu 2021:** "Graph neural networks, Geometric flows, and Neural diffusion PDEs". (200+ attendees)
- **Tutorial at SIGIR 2020:** "Searching the Web for Cross-lingual Parallel Data" [9] (500+ attendees virtual).
- **Tutorial at MLSS Summer School, Moscow 2019:** "Graph Neural Networks". (200+ attendees)
- **Tutorial at SIGMOD 2016:** "Automatic Entity Recognition and Typing in Massive Text Data" [19]. (200+ attendees).
- **Tutorial at NeurIPS 2017:** "Deep Learning on Graphs". (500+ attendees)
- **Tutorial at WWW 2016:** "Automatic Entity Recognition and Typing in Massive Text Corpora" [21]. (300+ attendees).
- **Tutorial at KDD 2015:** "Automatic Entity Recognition and Typing from Massive Text Corpora: A Phrase and Network Mining Approach" [20]. (300+ attendees).
- **Tutorial at KDD 2014:** "Bringing Structure to Text: Mining Phrases, Entities, Topics, and Hierarchies" [14] (300+ attendees).

REFERENCES

[1] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems* 26 (2013).

[2] Shaosheng Cao, Wei Lu, and Qiongkai Xu. 2015. Grarep: Learning graph representations with global structural information. In *Proceedings of the 24th ACM international on conference on information and knowledge management*. 891–900.

[3] Shaosheng Cao, Wei Lu, and Qiongkai Xu. 2016. Deep neural networks for learning graph representations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 30.

[4] Shiyu Chang, Wei Han, Jiliang Tang, Guo-Jun Qi, Charu C Aggarwal, and Thomas S Huang. 2015. Heterogeneous network embedding via deep architectures. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*. 119–128.

[5] Ting Chen and Yizhou Sun. 2017. Task-guided and path-augmented heterogeneous network embedding for author identification. In *Proceedings of the tenth ACM international conference on web search and data mining*. 295–304.

[6] H. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Aradhye, G. Anderson, G. Corrado, W. Chai, M. Ispir, et al. 2016. Wide & deep learning for recommender systems. In *DLRS*. 7–10.

[7] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM conference on recommender systems*. 191–198.

[8] Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. 2017. metapath2vec: Scalable representation learning for heterogeneous networks. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. 135–144.

[9] Ahmed El-Kishky, Philipp Koehn, and Holger Schwenk. 2020. Searching the Web for Cross-lingual Parallel Data. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2417–2420.

[10] Ahmed El-Kishky, Thomas Markovich, Kenny Leung, Frank Portman, Aria Haghighi, and Ying Xiao. 2022. kNN-Embed: Locally Smoothed Embedding Mixtures For Diverse Candidate Retrieval. *online* (2022).

[11] Ahmed El-Kishky, Thomas Markovich, Serim Park, Chetan Verma, Baekjin Kim, Ramy Eskander, Yury Malkov, Frank Portman, Sofía Samaniego, Ying Xiao, et al. 2022. TwHIN: Embedding the Twitter Heterogeneous Information Network for Personalized Recommendation. *arXiv preprint arXiv:2202.05387* (2022).

[12] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. 855–864.

[13] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. *Advances in neural information processing systems* 30 (2017).

[14] Jiawei Han, Chi Wang, and Ahmed El-Kishky. 2014. Bringing structure to text: mining phrases, entities, topics, and hierarchies. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. 1968–1968.

[15] Wang-Cheng Kang, Derek Zhiyuan Cheng, Tiansheng Yao, Xinyang Yi, Ting Chen, Lichan Hong, and Ed H Chi. 2020. Learning to embed categorical features without embedding tables for recommendation. *arXiv preprint arXiv:2010.10784* (2020).

[16] Maxim Naumov, Dheevatsa Mudigere, Hao-Jun Michael Shi, Jianyu Huang, Narayanan Sundaraman, Jongsoo Park, Xiaodong Wang, Udit Gupta, Carole-Jean Wu, Alisson G Azzolini, et al. 2019. Deep learning recommendation model for personalization and recommendation systems. *arXiv preprint arXiv:1906.00091* (2019).

[17] Aditya Pal, Chantat Eksombatchai, Yitong Zhou, Bo Zhao, Charles Rosenberg, and Jure Leskovec. 2020. Pinnersage: Multi-modal user embedding framework for recommendations at pinterest. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2311–2320.

[18] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. 701–710.

[19] Xiang Ren, Ahmed El-Kishky, Heng Ji, and Jiawei Han. 2016. Automatic entity recognition and typing in massive text data. In *Proceedings of the 2016 International Conference on Management of Data*. 2235–2239.

[20] Xiang Ren, Ahmed El-Kishky, Chi Wang, and Jiawei Han. 2015. Automatic Entity Recognition and Typing from Massive Text Corpora: A Phrase and Network Mining Approach. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2319–2320.

[21] Xiang Ren, Ahmed El-Kishky, Chi Wang, and Jiawei Han. 2016. Automatic entity recognition and typing in massive text corpora. In *Proceedings of the 25th International Conference Companion on World Wide Web*. 1025–1028.

[22] Emanuele Rossi, Fabrizio Frasca, Ben Chamberlain, Davide Eynard, Michael Bronstein, and Federico Monti. 2020. Sign: Scalable inception graph neural networks. *arXiv preprint arXiv:2004.11198* (2020).

[23] Jian Tang, Meng Qu, and Qiaozhu Mei. 2015. Pte: Predictive text embedding through large-scale heterogeneous text networks. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*. 1165–1174.

[24] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. 2015. Line: Large-scale information network embedding. In *Proceedings of the 24th international conference on world wide web*. 1067–1077.

[25] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *International conference on machine learning*. PMLR, 2071–2080.

[26] Daixin Wang, Peng Cui, and Wenwu Zhu. 2016. Structural deep network embedding. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. 1225–1234.

[27] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2014. Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575* (2014).

[28] Cheng Yang, Zhiyuan Liu, Deli Zhao, Maosong Sun, and Edward Chang. 2015. Network representation learning with rich text information. In *Twenty-fourth international joint conference on artificial intelligence*.

[29] Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L Hamilton, and Jure Leskovec. 2018. Graph convolutional neural networks for web-scale recommender systems. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 974–983.

[30] Rong Zhu, Kun Zhao, Hongxia Yang, Wei Lin, Chang Zhou, Baole Ai, Yong Li, and Jingren Zhou. 2019. Aligraph: a comprehensive graph neural network platform. *arXiv preprint arXiv:1902.08730* (2019).