

AliGraph: A Comprehensive Graph Neural Network Platform

Hongxia Yang
Alibaba Group
Hangzhou, China
yang.yhx@alibaba-inc.com

ABSTRACT

An increasing number of machine learning tasks require dealing with large graph datasets, which capture rich and complex relationship among potentially billions of elements. Graph Neural Network (GNN) becomes an effective way to address the graph learning problem by converting the graph data into a low dimensional space while keeping both the structural and property information to the maximum extent and constructing a neural network for training and referencing. However, it is challenging to provide an efficient graph storage and computation capabilities to facilitate GNN training and enable development of new GNN algorithms. In this paper, we present a comprehensive graph neural network system, namely AliGraph, which consists of distributed graph storage, optimized sampling operators and runtime to efficiently support not only existing popular GNNs but also a series of in-house developed ones for different scenarios. The system is currently deployed at Alibaba to support a variety of business scenarios, including product recommendation and personalized search at Alibaba's E-Commerce platform. By conducting extensive experiments on a real-world dataset with 492.90 million vertices, 6.82 billion edges and rich attributes, Ali-Graph performs an order of magnitude faster in terms of graph building (5 minutes vs hours reported from the state-of-the-art PowerGraph platform). At training, AliGraph runs 40%-50% faster with the novel caching strategy and demonstrates around 12 times speed up with the improved runtime. In addition, our in-house developed GNN models all showcase their statistically significant superiorities in terms of both effectiveness and efficiency (e.g., 4.12%–17.19% lift by F1 scores).

CCS Concepts/ACM Classifiers

Design and analysis of algorithms → Graph algorithms analysis

Author Keywords

Graph Neural Network; Large Scale; E-Commerce; Platform

BIOGRAPHY

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.

KDD '19, August 4–8, 2019, Anchorage, AK, USA.

© 2019 Copyright is held by the owner/author(s).

ACM ISBN 978-1-4503-6201-6/19/08.

DOI: <https://doi.org/10.1145/3292500.3340404>

Dr. Hongxia Yang is working as the Senior Staff Data Scientist and Director in Alibaba Group. Her interests span the areas of Bayesian statistics, time series analysis, spatial-temporal modeling, survival analysis, machine learning, data mining and their applications to problems in business analytics and big data. She used to work as the Principal Data Scientist at Yahoo! Inc and Research Staff Member at IBM T.J. Watson Research Center respectively and got her PhD degree in Statistics from Duke University in 2010. She has published over 40 top conference and journal papers and held 9 filed/to be filed US patents and is serving as the associate editor for Applied Stochastic Models in Business and Industry. She has been elected as Elected Members of the International Statistical Institute (ISI) in 2017 and the Chinese Institute of Electronics Young Scientist Club in 2019 respectively.



REFERENCES

- [1] Vincent, Z., Sha, M., Li, Y., Yang, H., Fang, Y., Zhang, Z. and Chang, K., Heterogeneous Embedding Propagation for Large-Scale E-Commerce User Alignment. IEEE International Conference on Data Mining series (ICDM), 2018.
- [2] Cen, Y., Zou, X., Zhang, J., Yang, H., Zhou, J. and Tang, J., Representation Learning for Attributed Multiplex Heterogeneous Network. 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2019.
- [3] Liu, N., Tan, Q., Li, Y., Yang, H., Zhou, J. and Hu, X., Is a Single Vector Enough? Exploring Node Polysemy for Network Embedding. 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2019.

- [4] Chen, Q., Lin, J., Zhang, Y., Yang, H., Zhou, J. and Tang, J., Towards Knowledge-Based Personalized Product Description Generation in E-commerce. 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2019.
- [5] Du, Z., Wang, X., Yang, H., Zhou, J. and Tang, J., Sequential Scenario-Specific Meta Learner for Online Recommendation. 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2019.
- [6] Zhu, R., Zhao, K., Yang, H., Lin, W., Zhou, C., Ai, B., Li, Y. and Zhou, J., AliGraph: A Comprehensive Graph Neural Network Platform. 45th International Conference on Very Large Data Bases (VLDB), 2019.
- [7] Zhao, Y., Wang, X., Yang, H., Song, L., Tang, J., Large Scale Evolving Graphs with Burst Detection. 28th International Joint Conference on Artificial Intelligence (IJCAI), 2019.
- [8] Li, C., Shen, D., Jia, K. and Yang, H., Hierarchical Representation Learning for Bipartite Graphs. 28th International Joint Conference on Artificial Intelligence (IJCAI), 2019.
- [9] Ding, M., Zhou, C., Chen, Q., Yang, H. and Tang, J., Cognitive Graph for Multi-Hop Reading Comprehension at Scale. 57th Annual Meeting of the Association for Computational Linguistics (ACL), 2019.