

Appendix

A On-line Performance Analysis

We extend our evaluation of the TESPA model to real-world industry scenarios, specifically conducting online experiments on a widely used sponsored search platform. Leveraging the learned item representations, we compute pairwise similarities, identifying top candidates for each item to construct a recommendation table. This table serves as a pivotal component supporting online services on the platform. The evaluation is carried out across two distinct industry scenarios: ProductAds and ProductReco. Key performance metrics such as Revenue Per Thousand (RPM), Click-Through-Rate (CTR), and Cost Per Click (CPC) are employed for a comprehensive assessment. The baseline model for comparison is BERT, a previously utilized online serving model. As depicted in Table 1, the TESPA model showcases a noteworthy enhancement in on-line performance. Particularly in the critical ProductAds scenario, TESPA demonstrates an 8.31% increase in RPM, a 9.25% improvement in CTR, and a 5.27% reduction in CPC. This signifies not only a considerable uplift in online revenue but also an enhancement in both user and advertiser experiences. The substantial improvements underscore the efficacy of our proposed model in real-world industry applications. It is worth noting that TESPA has seamlessly integrated into diverse industry scenarios, attesting to its versatility and applicability across different domains.

B Case Study

In this section, we showcase two illustrative cases retrieved from the online serving pipeline, as delineated in Table 2. The first case involves an input item, a backpack adorned with Ariel, a beloved cartoon character. Notably, the retrieved candidates exhibit a thematic congruence, featuring designs catered to a similar cartoon fashion aesthetic tailored for children. In the second case, the input item is a white bone china product, and the recalled candidates exhibit a discernible thematic affinity with the input item. Traditional models often rely on retrieving items solely based on semantic similarity, a practice that tends to result in low diversity and consequently, suboptimal user experiences. For instance, in the first case, traditional models might yield candidates comprising identical bags differing only in size. This limitation underscores the necessity for a more nuanced approach to retrieval. In contrast, the candidates retrieved by our proposed model, TESPA, demonstrate increased informativeness. This observation underscores the efficacy of the mutual reinforcement learning paradigm employed in TESPA. The model’s ability to discern and present more varied and informative candidates reflects its potential to significantly enhance the user experience by surpassing the limitations of conventional semantic similarity-based retrieval models.

C Related Work

Topological-based methods rely on topological features, especially the topological relationships introduced by users’ clicks and views. For topological-based methods, collaborative filtering [10, 16, 21, 22] is a typical technique that has been widely used; this approach

Table 1: Results of on-line test.

	RPM	CTR	CPC
ProductAds	+8.31%	+9.25%	-5.27%
ProductReco	+1.39%	+2.36%	-0.94%

Table 2: Case study of the on-line serving model.

Input Item	Candidate1	Candidate2
 <i>Title: Disney Ariel Toddler Backpack 12" Backpack for Kids</i>	 <i>Title: Disney Mickey Mouse Be Awesome Backpack School Kids</i>	 <i>Title: Nickelodeon Paw Patrol Boy's 18" Carry-On Duffel Bag</i>
 <i>Title: Noritake 40Pc Accompanist China Set, Service for 8</i>	 <i>Title: Noritake Blueshire China Soup Bowl Set Of 4</i>	 <i>Title: Noritake Blue Hill 2482 Pattern Service for 12</i>

jointly learns item and user’s latent embeddings to fit their interaction matrix. In recent years, conventional collaborative filtering has been enhanced by the deep neural networks [2, 8]. For such a regression-based task, matrix factorization (MF) methods associate each user and item with an embedding and model their matching score as the inner item between their embeddings, including Localized MF [28], Hierarchical MF [20] and Social-aware MF [29]. Meanwhile, by having items connected by their co-click relationships, the graph representation techniques are now being intensively applied to this area. In [4, 19], network embedding approaches are used for item representation; and more recently, graph neural networks are adopted [5, 23]. For example, SpectralCF [30] utilizes the spectral graph convolutions for collaborative filtering. GC-MC [17] and NGCF [7] adopt Graph Convolutional Network (GCN) to model user-item interactions. LR-GCCF [1], and LightGCN [[6] eliminate unnecessary non-linear operations, showing superior performance.

Apart from the above works, the utilization of textual feature is also important for item representation, especially in the text-rich scenarios. Many of the content based recommender systems leverage textual descriptions to formulate the fundamental representation of item [18, 24, 26]. Thanks to the quick development of pretrained language models [3, 12], much more powerful text encoders can be developed for the related applications [11, 14, 25], which significantly improves the representation quality.

Finally, the topological and textual features are also combined for better representation quality. A typical way of combination is demonstrated as the recent works PinSage [23] and TextGNN [33].

Given a graph of items connected by their co-click relations, each item is represented by its textual feature in the first place; then, it is aggregated with its neighbourhood embeddings via GNNs, which gives rise to the final item representation. HASH-CODE [27] developed a hierarchical contrastive learning architecture in different granularities, training the language model to refine the understanding of the text by graph topology and leveraging a high-frequency component (HFC)-aware contrastive learning objective that makes the learned embeddings more discriminative. Heterformer [9] proposed a network-empowered Transformer to inject heterogeneous structure information into each Transformer layer when encoding node texts, and designed a strategy for textless node learning with heterogeneous type-specific projection and embedding warm-up. Similar strategies are also applied in related scenarios, e.g., [13, 15, 31, 32].

Different from these methods, we focus on enriching the graph topology based on semantics and encoding the knowledge from the graph topology into the language models, and our proposed model employs a multi-channel co-training paradigm to preserve the fine-grained correlations.

References

- [1] Lei Chen, Le Wu, Richang Hong, Kun Zhang, and Meng Wang. 2020. Revisiting Graph Based Collaborative Filtering: A Linear Residual Graph Convolutional Network Approach. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7–12, 2020*. AAAI Press, 27–34. <https://ojs.aaai.org/index.php/AAAI/article/view/5330>
- [2] 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.
- [3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
- [4] Mihajlo Grbovic and Haibin Cheng. 2018. Real-time personalization using embeddings for search ranking at airbnb. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 311–320.
- [5] Will Hamilton, Zitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In *Advances in neural information processing systems*. 1024–1034.
- [6] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yong-Dong Zhang, and Meng Wang. 2020. LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25–30, 2020*, Jimmy X. Huang, Yi Chang, Xueqi Cheng, Jaap Kamps, Vanessa Murdock, Ji-Rong Wen, and Yiqun Liu (Eds.). ACM, 639–648. <https://doi.org/10.1145/3397271.3401063>
- [7] 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, WWW 2017, Perth, Australia, April 3–7, 2017*, Rick Barrett, Rick Cummings, Eugene Agichtein, and Evgeniy Gabrilovich (Eds.). ACM, 173–182. <https://doi.org/10.1145/3038912.3052569>
- [8] 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*. 173–182.
- [9] Bowen Jin, Yu Zhang, Qi Zhu, and Jiawei Han. 2023. Heterformer: Transformer-based deep node representation learning on heterogeneous text-rich networks. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 1020–1031.
- [10] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 42, 8 (2009), 30–37.
- [11] Rui Li, Jianan Zhao, Chaozhao Li, Di He, Yiqi Wang, Yuming Liu, Hao Sun, Senzhang Wang, Weiwei Deng, Yanming Shen, et al. 2022. HousE: Knowledge Graph Embedding with Householder Parameterization. *arXiv preprint arXiv:2202.07919* (2022).
- [12] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692* (2019).
- [13] Zhenghao Liu, Chenyan Xiong, Maosong Sun, and Zhiyuan Liu. 2019. Fine-grained Fact Verification with Kernel Graph Attention Network. *arXiv preprint arXiv:1910.09796* (2019).
- [14] Itzik Malkiel, Oren Barkan, Avi Caciularu, Noam Razin, Ori Katz, and Noam Koenigstein. 2020. RecoBERT: A Catalog Language Model for Text-Based Recommendations. *arXiv preprint arXiv:2009.13292* (2020).
- [15] Bochen Pang, Chaozhao Li, Yuming Liu, Jianxun Lian, Jianan Zhao, Hao Sun, Weiwei Deng, Xing Xie, and Qi Zhang. 2022. Improving Relevance Modeling via Heterogeneous Behavior Graph Learning in Bing Ads. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 3713–3721.
- [16] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. 2001. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*. 285–295.
- [17] Rianne van den Berg, Thomas N. Kipf, and Max Welling. 2017. Graph Convolutional Matrix Completion. *CoRR* abs/1706.02263 (2017). [arXiv:1706.02263](https://arxiv.org/abs/1706.02263)
- [18] 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*. 1235–1244.
- [19] Jizhe Wang, Pipei Huang, Huan Zhao, Zhibo Zhang, Binqiang Zhao, and Dik Lun Lee. 2018. Billion-scale commodity embedding for e-commerce recommendation in alibaba. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 839–848.
- [20] Suhang Wang, Jiliang Tang, Yilin Wang, and Huan Liu. 2015. Exploring Implicit Hierarchical Structures for Recommender Systems. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25–31, 2015*, Qiang Yang and Michael J. Wooldridge (Eds.). AAAI Press, 1813–1819. <http://ijcai.org/Abstract/15/258>
- [21] Yiqi Wang, Chaozhao Li, Mingzheng Li, Wei Jin, Yuming Liu, Hao Sun, Xing Xie, and Jiliang Tang. 2022. Localized Graph Collaborative Filtering. In *Proceedings of the 2022 SIAM International Conference on Data Mining (SDM)*. SIAM, 540–548.
- [22] Yiqi Wang, Chaozhao Li, Zheng Liu, Mingzheng Li, Jiliang Tang, Xing Xie, Lei Chen, and Philip S Yu. 2021. An Adaptive Graph Pre-training Framework for Localized Collaborative Filtering. *arXiv preprint arXiv:2112.07191* (2021).
- [23] 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.
- [24] Denghui Zhang, Yanchi Liu, Zixuan Yuan, Yanjie Fu, Haifeng Chen, and Hui Xiong. 2022. Multi-Faceted Knowledge-Driven Pre-Training for Product Representation Learning. *IEEE Transactions on Knowledge and Data Engineering* (2022).
- [25] Denghui Zhang, Zixuan Yuan, Yanchi Liu, Fuzhen Zhuang, and Hui Xiong. 2020. E-BERT: A Phrase and Product Knowledge Enhanced Language Model for E-commerce. *arXiv preprint arXiv:2009.02835* (2020).
- [26] Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. 2016. Collaborative knowledge base embedding for recommender systems. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 353–362.
- [27] Peiyang Zhang, Chaozhao Li, Liying Kang, Feiran Huang, Senzhang Wang, Xing Xie, and Sunghun Kim. 2024. High-Frequency-aware Hierarchical Contrastive Selective Coding for Representation Learning on Text Attributed Graphs. In *Proceedings of the ACM on Web Conference 2024*. 4316–4327.
- [28] Yongfeng Zhang, Min Zhang, Yiqun Liu, Shaoping Ma, and Shi Feng. 2013. Localized matrix factorization for recommendation based on matrix block diagonal forms. In *22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13–17, 2013*, Daniel Schwabe, Virgilio A. F. Almeida, Hartmut Glaser, Ricardo Baeza-Yates, and Sue B. Moon (Eds.). International World Wide Web Conferences Steering Committee / ACM, 1511–1520. <https://doi.org/10.1145/2488388.2488520>
- [29] Zhou Zhao, Hanqing Lu, Deng Cai, Xiaofei He, and Yueting Zhuang. 2016. User Preference Learning for Online Social Recommendation. *IEEE Trans. Knowl. Data Eng.* 28, 9 (2016), 2522–2534. <https://doi.org/10.1109/TKDE.2016.2569096>
- [30] Lei Zheng, Chun-Ta Lu, Fei Jiang, Jiawei Zhang, and Philip S. Yu. 2018. Spectral collaborative filtering. In *Proceedings of the 12th ACM Conference on Recommender Systems, RecSys 2018, Vancouver, BC, Canada, October 2–7, 2018*, Sole Pera, Michael D. Ekstrand, Xavier Amatriain, and John O'Donovan (Eds.). ACM, 311–319. <https://doi.org/10.1145/3240323.3240343>
- [31] Wanjun Zhong, Jingjing Xu, Duyu Tang, Zenan Xu, Nan Duan, Ming Zhou, Jiahai Wang, and Jian Yin. 2019. Reasoning over semantic-level graph for fact checking. *arXiv preprint arXiv:1909.03745* (2019).
- [32] Jie Zhou, Xu Han, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. 2019. GEAR: Graph-based evidence aggregating and reasoning for fact verification. *arXiv preprint arXiv:1908.01843* (2019).
- [33] Jason Zhu, Yanling Cui, Yuming Liu, Hao Sun, Xue Li, Markus Pelger, Liangjie Zhang, Tianqi Yan, Ruofei Zhang, and Huasha Zhao. 2021. TextGNN: Improving Text Encoder via Graph Neural Network in Sponsored Search. *arXiv preprint arXiv:2101.06323* (2021).