Complementary Product Recommendation for Long-tail Products

RASTISLAV PAPSO*, Kempelen Institute of Intelligent Technologies, Slovakia

Identifying complementary relations between products plays a key role in e-commerce Recommender Systems (RS). Existing methods in Complementary Product Recommendation (CPR), however, focus only on identifying complementary relations in huge and data-rich catalogs, while none of them considers real-world scenarios of small and medium e-commerce platforms with limited number of interactions. In this paper, we discuss our research proposal that addresses the problem of identifying complementary relations in such sparse settings. To overcome the data sparsity problem, we propose to first learn complementary relations in large and data-rich catalogs and then transfer learned knowledge to small and scarce ones. To be able to map individual products across different catalogs and thus transfer learned relations between them, we propose to create Product Universal Embedding Space (PUES) using textual and visual product meta-data, which serves as a common ground for the products from arbitrary catalog.

CCS Concepts: • **Information systems** → **Personalization**; **Electronic commerce**; • **Applied computing** → *Document management and text processing*.

Additional Key Words and Phrases: complementary products, personalization, e-commerce, latent embedding

ACM Reference Format:

Rastislav Papso. 2023. Complementary Product Recommendation for Long-tail Products. In *RecSys '23: The ACM Conference on Recommender Systems, September 18–23, 2023, Singapore.* ACM, New York, NY, USA, 10 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Since we have entered the information age, amount of consumable content on the internet started to grow exponentially. For e-commerce platforms, this poses a serious challenge to provide users with products of a high relevance to improve their customer experience. Although most of the work in RS focuses on understanding relations between users and products [12, 26], it is also important to understand relations between products themselves. In the e-commerce, there are two main types of relations - 1) substitute and 2) complementary [17]. Substitute products can be purchased *interchangeably* - for example two mobile phones of different type. On the other hand, complementary products are usually purchased *together* to serve a joint demand - for example a mobile phone and a mobile case. Complementary products are generally a basis for "cross-sell" and can significantly contribute to the revenue of the business.

While identifying substitute products is relatively straightforward (e.g., by obtaining the most similar products), finding complements poses more challenging task. At first, complementary products are not necessarily similar in the feature space [36]; using a naive distance measure thus may not work. Next, majority of work in CPR makes an assumption that complementary products are those bought together (co-purchased) [18, 27, 32, 35, 36]. This assumption, however, introduces non-negligible noise to the ground-truth variable, as the co-purchased products are not always in complementary relation [37]. Another problem that should be taken into account is the relation asymmetry [36]. For instance, it is likely that user would buy phone accessories such as battery charger or phone case while buying a phone,

*Also with Brno University of Technology Czech Republic.

but quite unlikely the other way around. In addition, by mining relations from existing co-purchases, one is able to find most popular complements, but cannot discover relations for a huge amount of products from long-tail (as there is not enough interaction with these). Finally, since the co-purchase strategy may work reasonably well on huge e-commerce platforms with immense number of purchases such as Amazon, Ebay or Taobao, it is not applicable to medium or small sized platforms. Since the number of purchases on the medium and small sized e-commerce platforms is considerably lower, there is much less informative signal available about complementary relations between products. In such sparse settings (i.e., limited data), standard methods (e.g., association rules [1, 37]) fail. This limitation poses a challenge for smaller platforms or their RS providers (recommendation-as-a-service (RaaS) businesses) to provide platform users with accurate complementary recommendations. It also leads to extending the gap and gaining market advantage of big e-commerce platforms over smaller ones.

One of the approaches that addresses problem with limited data and which became popular in recent years in various application is transfer learning [40]. Main idea of transfer learning is to train a general-purpose model on vast amount of data, and then apply (i.e., transfer) learned knowledge on a more specific domain or task, in which little to no data is available. Transfer learning was successfully applied to numbers of tasks in Computer Vision and Natural Language Processing lately [22, 25]. Utilizing its idea in RS is, however, more tricky. Since the majority of approaches in RS work with catalog specific information about products (e.g., product IDs, product categories), learned knowledge is bound to the catalog and thus not transferable. Cross-domain recommendation attempts to apply transfer learning in RS settings [38]. However, most of the methods in cross-domain recommendation assume known overlap of users and/or products in considered domains or catalogs, which is usually not the case in real-world scenarios.

In our work, we systematically investigate main challenges in the CPR, such as relation asymmetry, determination of ground truth labels and discovering missing relations across products. Our research aim is focused on applicability of proposed methods to real-world scenarios of medium and small sized e-commerce platforms, i.e., scenarios with limited data - translated to research objectives as follows:

- **RO1: Identifying complementary relations.** Besides mining product relations through existing co-purchases, it is vital to focus on discovering new (missing) relations. For this, we propose to consider asymmetric nature of complements and noisy ground truth labels together with the additional information on products meta-data and knowledge extracted from LLMs. Complementary relation identification is performed on large and data-rich catalogs (addressed by RQ2, RQ3).
- **RO2: Knowledge transfer to sparse catalogs.** After identifying complementary relations in data-rich settings, the acquired knowledge can be applied to sparse catalogs with little to no co-purchase data. To be able to transfer this knowledge, we propose to create product unifying embedding space computed from product textual and visual meta-data optimized for the e-commerce domain (addressed by RQ1, RQ4).

2 RELATED WORK

2.1 Product representation

To overcome the well known cold-start problem, authors in RS started to use product representations obtained from textual or visual product meta-data. With such representations, one is able to learn relations on a higher level of abstraction. For instance, on the IDs level we can extract the knowledge tight to product IDs, e.g., *ID 1* is co-purchased with *ID 2*. On the product meta-data, we are able to learn relations at higher level of abstraction, such as *red dress* is frequently co-purchased with *red heels*, or in the PUES, a part of the latent space is co-purchased with another part of the latent space (i.e., all products that represent red dresses and red heels are complements with a high probability).

Representation of products based on product meta-data is usually obtained by topic modelling [18] or, more recently, by using pre-trained deep learning models operating with text [2, 7] or vision [10, 35]. While using pre-trained deep learning models is a reasonable choice, these models are usually trained on general-purpose datasets [4, 8]. By fine-tuning the models on specific domain, their performance can be significantly increased [16]. Moreover, since the models are trained in a way to obtain general knowledge of language or vision itself, e-commerce data will form only a small portion of the final embedding space [34], which can cause representation degeneration and thus poor performance on subsequent tasks [5, 29], including the complementary product recommendation. In addition, most works in CPR use only one modality (textual or visual) at the time, whereas each modality can have varying level of importance across individual domains in e-commerce. Finally, information from various modalities can reduce the sparsity problem which differs across the e-commerce segments (i.e., it is rare to have rich textual description for fashion products, however, it is likely that various images will be available).

To obtain effective product representation that will serve as a basis for complementary relation mining, we propose to create Product Universal Embedding Space based on the multi-modal nature of the products, i.e., textual and visual representation, into one common embedding space. Textual and visual modality shall be combined together to solve segment-specific data sparsity problem. Embeddings shall be fine-tuned on e-commerce domain to adapt them to the domain and thus overcome the representation degeneration problem and to improve the performance on the subsequent tasks (i.e., by using them in CPR).

2.2 Complementary Product Recommendation

Main goal of CPR is to mine complementary relations between products from the data. The most straightforward way how to discover such relations is utilizing association rules on co-purchases [1, 37]. While association rules could be used on dense catalogs with relatively rich history of purchases, they fail in sparse settings with not enough co-purchase data. In addition, association rules are able to reveal only existing relations (i.e., frequently co-purchased products), and cannot discover new ones that can be missing in the data.

More recent approaches deal with the CPR problem by modelling product relations with knowledge graphs [7, 18, 31, 36]. Products are represented as nodes and relations between them are represented as edges. There are typically more types of relations, while most considered are substitute and complementary. Such approaches not only derive existing relations, but also try to predict new relations (i.e., link prediction task [13]). Historically, shallow models were applied such as logistic regression [18] or skip-gram-like models [30, 36] on product representations for the link prediction problem. Recent advances in the field of Graph Neural Networks (GNN) [6, 28] inspired authors to apply GNNs in CPR domain, representing nodes (products) as low-dimensional vectors while preserving both network topology structure and product features and easily perform graph-related tasks [7, 14, 15, 31]. Graph-based models often work with product representations derived from textual [7, 18, 31] or visual [9, 19, 27] meta-data and thus overcome the activity related cold-start problem. However, the extracted knowledge (i.e., complements) is catalog specific and thus not transferable to entirely new catalogs. The main reason is that these models use catalog specific meta-information about the products, such as category label or product type to reduce search space in the process of finding the complements [7, 18, 30, 31]. The knowledge they obtain is hence bounded to the specific data and its category taxonomy.

In our work we focus on learning complementary relations in a way that makes the learned knowledge applicable to a long-tail products or new catalog. For this purpose, using the Product Universal Embedding Space (PUES), and utilizing additional knowledge extracted by LLMs prompting, the graph-based link prediction is used to reveal missing complementary relations. In the next step, the PUES allows us to transfer the learned knowledge to new products or new catalogs.

3 RESEARCH PROPOSAL

Our research proposal tackles with well-known problem in RS domain - complementary product recommendations. Specifically, we aim to improve the first and mandatory step - a complementary relations mining. State-of-the-art approaches heavily assume that complementary products are often co-purchased, which does not hold for many examples. Despite that, there is a significant number of long-tail products, where standard approaches obviously fail, as there is no or limited amount of user activity. Hand by hand with the importance of identified research challenges, our proposal has a direct industry application by providing a solution for small businesses or their RaaS providers to Complementary Product Recommendation (CPR). Whereas none of the methods in the CPR considers sparse and data-limited settings of such businesses, our research fills this gap.

Based on the state-of-the-art analysis, we have identified several open research problems translated to the research objectives (described above) which we address in our proposal. Our approach consists of three major steps (Figure 1), which each separately improves particular tasks in CPR. As a first step, the product catalog is represented in the Product Universal Embedding Space. We employ textual and visual product meta-data and fine-tune pre-trained deep learning models for the product similarity task. In the next phase, the PUES is used as a basis for complementary relations mining. However, as there is significant amount of long-tail products, these relations are enhanced with the graph-based link prediction and with the additional knowledge extracted via the LLMs prompting. As a result, we obtain an oriented graph representation of complementary relations for a given catalog. Finally, when a new product or even a new catalog appears, we employ the transfer learning to project new products into the PUES and to impute the complementary relations based on the pre-learned knowledge. To further explore each step, we have defined four Research Questions (RQs) we will further explore.

RQ1: Can Product Universal Embedding Space learned on various e-commerce datasets (and domains) improve performance of subsequent tasks, specifically in highly dynamic scenarios of continuous appearance of new products?

To construct PUES, we propose to utilize deep learning models for text and vision. To obtain better representation of products and thus improve the performance on subsequent tasks and reduce the representation degeneration problem, we will fine-tune these pre-trained models on the e-commerce data. In other words, we will fine-tune the models to obtain embedding space, in which two similar products should be represented close to each other. We will specifically focus on notoriously problematic cases in e-commerce when almost identical products (based on the text description or image), e.g., *men black t-shirt* vs. *women black t-shirt*, are represented close to each other. This is a sub-optimal representation which causes a serious problem for real-world applications. Besides measuring embedding quality by employing them on subsequent tasks [20], recent work suggested that quality of embedding space can be also measured by *alignment* and *uniformity*, and theoretically and empirically showed that employing self-supervised contrastive learning [3] directly optimizes for these two properties [29]. Therefore in our work, we use contrastive learning to fine-tune vision and text deep learning models for better adaptation to e-commerce domain and thus obtaining product representations with desired properties.

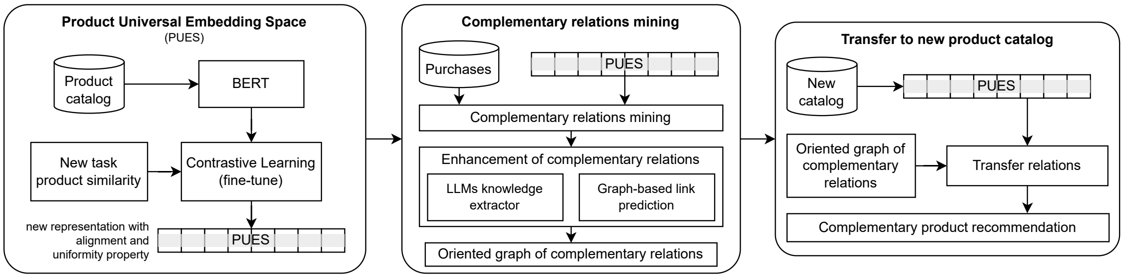

Fig. 1. Proposed research idea consists of three major components (left to right): Product Universal Embedding Space (PUES) represents products across various categories in one universal space (guaranteeing alignment and uniformity property); next PUES together with user purchases is used to mine complementary relations, which are further extended based on the LLMs knowledge and graph link prediction; finally the transfer learning is used for new catalogs or domains to generate complementary products.

RQ2: How the application of Large Language Models for the knowledge extraction on complementary relations improves the complement identification tasks in long-tail products scenario?

Complementary relation between two products is often determined by their co-purchase frequency. However, frequently co-purchased products are not necessarily complementary [37]. In CPR, majority of the work does not address this problem and assume that co-purchased products are in complementary relation. We argue that this is far from the truth, and specifically in medium size e-commerce stores, there is a lot of bias caused by random customers (i.e., the visit and purchase is driven by the low price product search, rather than regular shopping behavior). To address this problem, we explore possibility of employing prompt-based Large Language Models (LLMs), such as currently popular ChatGPT, for ground truth label generation. Results of recent work [39] indicate that LLMs do have the potential to handle data annotation tasks (via the text-based prompting). Based on the preliminary explorations, there are domains where the high level knowledge can be extracted. Further experiments with different types of prompts need to be conducted to determine optimal and universal prompting strategy. For instance we can ask for complements to the product given product title or the higher level knowledge can be extracted on the product sub-category levels, e.g., "What is the complement to the PC monitor?" We believe that such additional knowledge can notably improve the complementary relations identification, especially for niche domains, where there is a low user activity, or low number of multiple product purchases available.

RQ3: Is it beneficial to apply graph-based model for modeling the relation asymmetry and to discover missing complementary relations for the complementary products identification task?

When dealing with complementary relations learning, one has to bear in mind that these relations are often asymmetric. A typical example is that users often buy chargers for their phones however they rarely buy phones for their chargers [32]. When using graph-based model to learn complementary relations between products, a natural choice how to model asymmetry is by modelling relations as directed edges. Therefore in our work, we will model products and their complementary relations by a directed graph. Modelling CPR problem by graph comes with several advantages, such as ability to easily perform graph-based algorithms, e.g., algorithms for link prediction [13], which is useful for discovering missing relations in the graph (and thus improve the CPR for long-tail products).

Based on the standard co-purchasing definition, there is a significant amount of products for which no complements exist. In other words, even if we have enough complements identified either from user activity or from data annotation, there are still particular relations that are missing. This is especially true for new or long-tail products, for which finding complements can be challenging. Previous works in CPR that approached revealing missing relations by a link prediction reported promising results [18, 30, 36]. However in these works, link prediction algorithms that do not consider current progress in link prediction domain, were used. We will therefore experiment with application of several state-of-the-art link prediction algorithms, such as SEAL [33] or SimplE [11], to discover missing relations in the product graph and utilize the additional information about the products represented in the PUES (researched in RQ1).

RQ4 Are the complementary relations discovered via the proposed approach transferable to unseen and sparse catalogs?

Mining complementary relations from huge catalogs with vast amount of co-purchase data is challenging yet still achievable task. The same approaches are, however, not applicable to new catalogs with limited or no co-purchase data. Still an extensive human power is needed for manual annotation, which is obviously a sub-optimal and not scalable approach. Therefore we propose to apply ideas from transfer learning [40] and transfer learned knowledge from dense to sparse catalogs. The problem with knowledge transfer between catalogs lies in inability to map individual products across the catalogs (different *IDs* or slightly different portfolio). We believe that proposed PUES, which will serve as an unified embedding space for products from arbitrary catalog, will enable such unique application. Then, by learning complementary relations in this space, learned knowledge will be transferable to new catalogs.

4 PRELIMINARY RESULTS

The effective and transferable representation of products in the PUES is a core component of our research proposal. To gain evidence for formulating the research questions and check our initial hypothesis we conduct a series of experiments.

4.1 Methodology

Common approach to determine and to evaluate the properties of embedding space is to apply them in subsequent task(s), i.e., indirect evaluation [20]. To find out if we can obtain meaningful representation of products from their textual meta-data, we employ Sentence-BERT [24] to compute the product embeddings (it will also serve as one of the baselines for RQ1). To compute the embeddings, we utilize all available textual meta-data of the product, such as title, description, brand, technical details, etc. All of the fields are concatenated to single string and put to the Sentence-BERT as input. Output of the model is d-dimensional vector representation of the input text.

For the subsequent task, we opt for the state-of-the-art next-item sequence recommender CARCA [23], published at RecSys22. CARCA utilizes attributes of the products as additional information. The source-code for out implementation is publicly available[1]. Following the original experiments, we used 3 different versions of CARCA. First version used only product IDs as input, second used product IDs along with one-hot encoded categorical features of products (e.g., fine-grained category, brand) and third used IDs with textual embeddings computed by Sentence-BERT.

Data. The experiments were conducted on widely used Amazon Reviews Dataset [21]. Dataset contains subsets for each of 29 main product categories. Each subset is available in two versions - 1) full version and 2) five-core version in which all users and products have at least 5 interactions. For our experiments, we picked 10 datasets with varying sparsity, and numbers of products and users (for data statistics see Table 1). For the Video Games subset, we created

[1] http://www.anonymized.an

Table 1. Dataset statistics. *Full* is version with all users and products. *5core* contains only users and products that have at least 5 actions. *Sub* is artificially reduced *Full* version.

#	Dataset	Version	Users	Products	Actions	Avg. actions/user	Avg. actions/product
1	Arts, Crafts and Sewing	5core	56,210	22,855	492,492	8.76	21.55
2	Arts, Crafts and Sewing	Full	1,576,189	302,372	2,730,361	1.73	9.03
3	Cell Phones	5core	157,212	48,172	1,128,267	7.18	23.42
4	Industrial and Scientific	Full	1,245,370	165,682	1,711,934	1.37	10.33
5	Musical Instruments	5core	27,530	10,611	231,312	8.40	21.80
6	Musical Instruments	Full	903,050	112,128	1,472,323	1.63	13.13
7	Prime Pantry	Full	247,640	10,812	448,035	1.89	41.44
8	Video Games	5core	55,223	17,389	496,315	8.99	28.54
9	Video Games	Full	1,539,731	71,908	2,487,867	1.62	34.60
10	Video Games	Sub	597,496	55,734	765,738	1.28	13.74

also *sub* version, which is artificially reduced version to lower the average actions per product. It serves to evaluate the hypothesis that textual embeddings notably improve model performance in cases with small no. of actions per product.

Training and evaluation. Widely adopted leave-one-out protocol was used to evaluate the model [23] - last interaction from each sequence of the user's history is left out. The model is trained on the rest of the users' history. Then, evaluation is performed by sampling 100 negative products that were not interacted with by the user, and rank the positive (left-out) product among them. Finally, top 10 ranked products are selected for each user and average Hit-Rate (HR) and Normalized Discounted Cumulative Gain (NDCG) are computed. For more details, we refer reader to original paper [23], as we used the same training and evaluation methodology.

Hyperparameters. We followed suggestions in the original paper of CARCA method [23] - we used fixed hyper-parameter setup for each dataset. For the architecture of CARCA, we used three attention blocks. Number of heads per attention layer was set to 3. We used Adam optimizer with learning rate set to 0.001. Dropout rate was set to 0.5. Dimension d was set to 90, g was set to 450. L2 regularization was set to 0. We trained the model for 500 epochs with early stopping if no improvement was recorded during 20 subsequent epochs. Length of the sequence was set to 50.

4.2 Results

Results show that in some cases (i.e., dataset characteristics), incorporating textual embeddings as input for next-item recommendation significantly improves both HR and NDCG metrics. Hence in some cases, using the embeddings does not lead to considerable improvements (Table 2). Despite that the embedding version of the method outperforms other versions in the HR metric, surprisingly in five cases achieves sub-optimal performance in NDCG (which takes in the account the quality of recommended list order).

We have further explored various dataset characteristics, in order to identify a common properties, which can indicate the reason why improvements vary. Based on the comparison it seems that datasets with low *avg. number of actions per product* benefit from the textual embeddings the most. On the contrary, when a sufficient activity is available, the sequences of product IDs sufficiently describe the user behavior and embeddings are not so beneficial. It also may illustrate, that the standard text-based embedding (which are not fine-tuned for the e-commerce application) include a lot of bias which degrade the performance (short text description with specific language different from natural texts, i.e., product names). This supports our research objective and research questions that while using textual embeddings

Table 2. Experimental results. While in #1, #3, #5, #7, #8 and #9, the performance of ID+SBERT recorded slight or no performance gain (up to +2%) relative to ID version; in #2, #4, #6 and #10, the performance gain was considerably higher (up to +11%). According to Table 1, these datasets have the *lowest* avg. actions/product, which means that in such cases, textual embeddings have notable positive influence on the model performance.

#	Dataset	Version	CARCA ID		CARCA ID+onehot		CARCA ID+SBERT	
			HR@10	NDCG@10	HR@10	NDCG@10	HR@10	NDCG@10
1	Arts, Crafts and Sewing	5core	0.6589	**0.4813**	0.6570	0.4416	**0.6717**	0.4734
2	Arts, Crafts and Sewing	Full	0.6702	0.4661	0.7651	0.5645	**0.7742**	**0.5799**
3	Cell Phones	5core	0.6042	0.4009	0.6200	**0.4111**	**0.6247**	0.4094
4	Industrial and Scientific	Full	0.6472	0.4310	0.6527	0.4392	**0.6780**	**0.4814**
5	Musical Instruments	5core	0.5857	**0.3905**	0.5709	0.3645	**0.5863**	0.3783
6	Musical Instruments	Full	0.6912	0.4755	0.7301	0.5107	**0.7697**	**0.5659**
7	Prime Pantry	Full	0.5432	0.3077	0.5350	0.3058	**0.5594**	**0.3196**
8	Video Games	5core	0.7268	**0.5094**	0.7188	0.4861	**0.7278**	0.4964
9	Video Games	Full	0.8261	**0.6309**	0.8224	0.5896	**0.8404**	0.6279
10	Video Games	Sub	0.5726	0.3535	0.5787	0.3642	**0.6546**	**0.4198**

of products makes sense and can bring some performance gain, there is a space for improvement of embeddings themselves.

5 CONCLUSIONS

In this paper, we discussed our research proposal to Complementary Product Recommendation for catalogs with limited data. For modelling complementary relations between products, knowledge graphs are natural choice. There are several advantages of using graph-based models, such as handling relation asymmetry with ease by using directed edges or employing graph-based algorithms, namely link prediction for discovery of missing relations. Majority of work in CPR considers co-purchased products as complementary. This is however not always the case, since people regularly buy unrelated products together. To determine accurate ground truth complements, we propose to employ prompt-based LLMs like ChatGPT to perform automated data annotation. To address low resource (i.e., sparse) settings of medium and small e-commerce platforms or their RS providers (RaaS businesses), we propose to apply idea of transfer learning. To overcome the problem with mapping products across catalogs, we create Product Universal Embedding Space, which projects products from arbitrary catalogs to one, unified embedding space. Then, by learning the complementary relations in this space, we are able to transfer learned knowledge across catalogs and long-tail products. To construct the embedding space, we use product textual and visual meta-data and deep learning models. These models are fine-tuned on e-commerce data to improve quality of the embeddings and overcome the representation degeneration problem. In addition, such space is not only viable for CPR, but can serve as a basis to several subsequent tasks in RS, including product clustering, product categorization, product similarity search and others.

ACKNOWLEDGMENTS

I would like to give a special thanks to my supervisor XXXX for his continuous support, feedback, and encouragement. The work is conducted as a collaboration with the XXXX.

REFERENCES

[1] Rakesh Agrawal and Ramakrishnan Srikant. 1994. Fast Algorithms for Mining Association Rules in Large Databases. In *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB '94)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 487–499.

[2] Marina Angelovska, Sina Sheikholeslami, Bas Dunn, and Amir H. Payberah. 2021. Siamese Neural Networks for Detecting Complementary Products. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop*. Association for Computational Linguistics, Online, 65–70. https://doi.org/10.18653/v1/2021.eacl-srw.10

[3] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A Simple Framework for Contrastive Learning of Visual Representations. In *Proceedings of the 37th International Conference on Machine Learning (ICML'20)*. JMLR.org, Article 149, 11 pages.

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. https://doi.org/10.18653/v1/N19-1423

[5] Jun Gao, Di He, Xu Tan, Tao Qin, Liwei Wang, and Tie-Yan Liu. 2019. Representation Degeneration Problem in Training Natural Language Generation Models. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net. https://openreview.net/forum?id=SkEYojRqtm

[6] William L. Hamilton, Rex Ying, and Jure Leskovec. 2017. Representation Learning on Graphs: Methods and Applications. *IEEE Data Eng. Bull.* 40, 3 (2017), 52–74. http://sites.computer.org/debull/A17sept/p52.pdf

[7] Junheng Hao, Tong Zhao, Jin Li, Xin Luna Dong, Christos Faloutsos, Yizhou Sun, and Wei Wang. 2020. P-Companion: A Principled Framework for Diversified Complementary Product Recommendation. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* (Virtual Event, Ireland) *(CIKM '20)*. Association for Computing Machinery, New York, NY, USA, 2517–2524. https://doi.org/10.1145/3340531.3412732

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 770–778. https://doi.org/10.1109/CVPR.2016.90

[9] Ruining He, Charles Packer, and Julian McAuley. 2016. Learning compatibility across categories for heterogeneous item recommendation. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*. IEEE, 937–942.

[10] Wang-Cheng Kang, Eric Kim, Jure Leskovec, Charles Rosenberg, and Julian J. McAuley. 2018. Complete the Look: Scene-based Complementary Product Recommendation. *CoRR* abs/1812.01748 (2018). arXiv:1812.01748 http://arxiv.org/abs/1812.01748

[11] Seyed Mehran Kazemi and David Poole. 2018. SimplE Embedding for Link Prediction in Knowledge Graphs. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems* (Montréal, Canada) *(NIPS'18)*. Curran Associates Inc., Red Hook, NY, USA, 4289–4300.

[12] Yehuda Koren, Steffen Rendle, and Robert Bell. 2021. Advances in collaborative filtering. *Recommender systems handbook* (2021), 91–142.

[13] Ajay Kumar, Shashank Sheshar Singh, Kuldeep Singh, and Bhaskar Biswas. 2020. Link prediction techniques, applications, and performance: A survey. *Physica A: Statistical Mechanics and its Applications* 553 (2020), 124289. https://doi.org/10.1016/j.physa.2020.124289

[14] Weiwen Liu, Yin Zhang, Jianling Wang, Yun He, James Caverlee, Patrick P. K. Chan, Daniel S. Yeung, and Pheng-Ann Heng. 2022. Item Relationship Graph Neural Networks for E-Commerce. *IEEE Transactions on Neural Networks and Learning Systems* 33, 9 (2022), 4785–4799. https://doi.org/10.1109/TNNLS.2021.3060872

[15] Yiding Liu, Yulong Gu, Zhuoye Ding, Junchao Gao, Ziyi Guo, Yongjun Bao, and Weipeng Yan. 2020. Decoupled Graph Convolution Network for Inferring Substitutable and Complementary Items. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* (Virtual Event, Ireland) *(CIKM '20)*. Association for Computing Machinery, New York, NY, USA, 2621–2628. https://doi.org/10.1145/3340531.3412695

[16] Itzik Malkiel, Oren Barkan, Avi Caciularu, Noam Razin, Ori Katz, and Noam Koenigstein. 2020. RecoBERT: A Catalog Language Model for Text-Based Recommendations. *CoRR* abs/2009.13292 (2020). arXiv:2009.13292 https://arxiv.org/abs/2009.13292

[17] Andreu Mas-Colell, Michael D. Whinston, and Jerry R. Green. 1995. *Microeconomic Theory.* Oxford University Press, New York.

[18] Julian McAuley, Rahul Pandey, and Jure Leskovec. 2015. Inferring Networks of Substitutable and Complementary Products. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Sydney, NSW, Australia) *(KDD '15)*. Association for Computing Machinery, New York, NY, USA, 785–794. https://doi.org/10.1145/2783258.2783381

[19] Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton van den Hengel. 2015. Image-Based Recommendations on Styles and Substitutes. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Santiago, Chile) *(SIGIR '15)*. Association for Computing Machinery, New York, NY, USA, 43–52. https://doi.org/10.1145/2766462.2767755

[20] Niklas Muennighoff, Nouamane Tazi, Loic Magne, and Nils Reimers. 2023. MTEB: Massive Text Embedding Benchmark. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, Dubrovnik, Croatia, 2014–2037. https://aclanthology.org/2023.eacl-main.148

[21] Jianmo Ni, Jiacheng Li, and Julian McAuley. 2019. Justifying Recommendations using Distantly-Labeled Reviews and Fine-Grained Aspects. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, Hong Kong, China, 188–197. https://doi.org/10.18653/v1/D19-1018

[22] Yifan Peng, Shankai Yan, and Zhiyong Lu. 2019. Transfer Learning in Biomedical Natural Language Processing: An Evaluation of BERT and ELMo on Ten Benchmarking Datasets. In *Proceedings of the 18th BioNLP Workshop and Shared Task*. Association for Computational Linguistics, Florence,

Italy, 58–65. https://doi.org/10.18653/v1/W19-5006

[23] Ahmed Rashed, Shereen Elsayed, and Lars Schmidt-Thieme. 2022. Context and Attribute-Aware Sequential Recommendation via Cross-Attention. In *Proceedings of the 16th ACM Conference on Recommender Systems* (Seattle, WA, USA) *(RecSys '22)*. Association for Computing Machinery, New York, NY, USA, 71–80. https://doi.org/10.1145/3523227.3546777

[24] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, Hong Kong, China, 3982–3992. https://doi.org/10.18653/v1/D19-1410

[25] Manali Shaha and Meenakshi Pawar. 2018. Transfer Learning for Image Classification. In *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*. 656–660. https://doi.org/10.1109/ICECA.2018.8474802

[26] Xiaoyuan Su and Taghi M Khoshgoftaar. 2009. A survey of collaborative filtering techniques. *Advances in artificial intelligence* 2009 (2009).

[27] A. Veit, B. Kovacs, S. Bell, J. McAuley, K. Bala, and S. Belongie. 2015. Learning Visual Clothing Style with Heterogeneous Dyadic Co-Occurrences. In *2015 IEEE International Conference on Computer Vision (ICCV)*. IEEE Computer Society, Los Alamitos, CA, USA, 4642–4650. https://doi.org/10.1109/ICCV.2015.527

[28] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net. https://openreview.net/forum?id=rJXMpikCZ

[29] Tongzhou Wang and Phillip Isola. 2020. Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere. In *Proceedings of the 37th International Conference on Machine Learning (ICML'20)*. JMLR.org, Article 921, 11 pages.

[30] Zihan Wang, Ziheng Jiang, Zhaochun Ren, Jiliang Tang, and Dawei Yin. 2018. A Path-Constrained Framework for Discriminating Substitutable and Complementary Products in E-Commerce. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining* (Marina Del Rey, CA, USA) *(WSDM '18)*. Association for Computing Machinery, New York, NY, USA, 619–627. https://doi.org/10.1145/3159652.3159710

[31] An Yan, Chaosheng Dong, Yan Gao, Jinmiao Fu, Tong Zhao, Yi Sun, and Julian Mcauley. 2022. Personalized Complementary Product Recommendation. In *Companion Proceedings of the Web Conference 2022* (Virtual Event, Lyon, France) *(WWW '22)*. Association for Computing Machinery, New York, NY, USA, 146–151. https://doi.org/10.1145/3487553.3524222

[32] Hang Yu, Lester Litchfield, Thomas Kernreiter, Seamus Jolly, and Kathryn Hempstalk. 2019. Complementary Recommendations: A Brief Survey. In *2019 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS)*. 73–78. https://doi.org/10.1109/HPBDIS.2019.8735479

[33] Muhan Zhang and Yixin Chen. 2018. Link Prediction Based on Graph Neural Networks. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems* (Montréal, Canada) *(NIPS'18)*. Curran Associates Inc., Red Hook, NY, USA, 5171–5181.

[34] Shu Zhang, Ran Xu, Caiming Xiong, and Chetan Ramaiah. 2022. Use All the Labels: A Hierarchical Multi-Label Contrastive Learning Framework. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 16660–16669.

[35] Yin Zhang, Haokai Lu, Wei Niu, and James Caverlee. 2018. Quality-Aware Neural Complementary Item Recommendation. In *Proceedings of the 12th ACM Conference on Recommender Systems* (Vancouver, British Columbia, Canada) *(RecSys '18)*. Association for Computing Machinery, New York, NY, USA, 77–85. https://doi.org/10.1145/3240323.3240368

[36] Tong Zhao, Julian McAuley, Mengya Li, and Irwin King. 2017. Improving recommendation accuracy using networks of substitutable and complementary products. In *2017 International Joint Conference on Neural Networks (IJCNN)*. 3649–3655. https://doi.org/10.1109/IJCNN.2017.7966315

[37] Jiaqian Zheng, Xiaoyuan Wu, Junyu Niu, and Alvaro Bolivar. 2009. Substitutes or Complements: Another Step Forward in Recommendations. In *Proceedings of the 10th ACM Conference on Electronic Commerce* (Stanford, California, USA) *(EC '09)*. Association for Computing Machinery, New York, NY, USA, 139–146. https://doi.org/10.1145/1566374.1566394

[38] Feng Zhu, Yan Wang, Chaochao Chen, Jun Zhou, Longfei Li, and Guanfeng Liu. 2021. Cross-Domain Recommendation: Challenges, Progress, and Prospects. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, Zhi-Hua Zhou (Ed.). International Joint Conferences on Artificial Intelligence Organization, 4721–4728. https://doi.org/10.24963/ijcai.2021/639 Survey Track.

[39] Yiming Zhu, Peixian Zhang, Ehsan-Ul Haq, Pan Hui, and Gareth Tyson. 2023. Can ChatGPT Reproduce Human-Generated Labels? A Study of Social Computing Tasks. *arXiv preprint arXiv:2304.10145* (2023).

[40] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. 2020. A comprehensive survey on transfer learning. *Proc. IEEE* 109, 1 (2020), 43–76.

Received 23 May 2023