VIER: Visual Imagination Enhanced Retrieval in Sponsored Search

Yadong Zhang
Meituan
Beijing, China
zhangyadong05@meituan.com

Yuqing Song
Meituan
Beijing, China
songyuqing03@meituan.com

Siyu Lu
Meituan
Beijing, China
lusiyu05@meituan.com

Qiang Liu
Meituan
Beijing, China
liuqiang43@meituan.com

Xingxing Wang
Meituan
Beijing, China
wangxingxing04@meituan.com

Abstract

Embedding-based Retrieval (EBR) has been a fundamental component in sponsored-search systems, which retrieves high-quality products for the user's search query by encoding the information of the query, user and product into dense embeddings. However, due to the characteristic of location-based service, the user input queries suffer from two extremes: *overly brief queries with vague intentions and lengthy queries with substantial noise*, both of which make it challenging to discern the exact user search intent. In fact, the e-consumers typically have a mental imagery of the product they intend to search for, reflecting their specific purchasing intentions. In this paper, we propose a Visual Imagination Enhanced Retrieval model (VIER) to explore the implicit imagery of users. Specifically, we design a visual imagination network to reconstruct the imagery embeddings that capture both coarse-grained query commonalities and fine-grained user personalities. These pseudo-image representations are integrated with the query and user behavior to enhance the understanding of user search intentions for improved retrieval. According to online A/B tests on Meituan sponsored-search system, our method significantly outperforms baselines in terms of revenue, clicks and click-through rate.

CCS Concepts

• **Information systems** → **Information retrieval**.

Keywords

sponsored search, dense retrieval, visual imagination

ACM Reference Format:
Yadong Zhang, Yuqing Song, Siyu Lu, Qiang Liu, and Xingxing Wang. 2024. VIER: Visual Imagination Enhanced Retrieval in Sponsored Search. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24), October 21–25, 2024, Boise, ID, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3627673.3680005

CIKM '24, October 21–25, 2024, Boise, ID, USA

ACM ISBN 979-8-4007-0436-9/24/10
https://doi.org/10.1145/3627673.3680005

1 Introduction

Embedding-based Retrieval (EBR) has become an important component of sponsored-search systems across platforms such as Meituan, Facebook, Taobao and more [8–10, 16, 17]. It encodes the information of queries, users, and ad products into dense embeddings to retrieve high-quality products for user search queries. However, in the location based service, the users' search queries usually fall into two extremes: 1) overly brief queries with vague intentions, such as "flower" and "fruit", lack specific user preference expressions, making it challenging to retrieve products that users prefer. 2) Lengthy queries with substantial noise that combine multiple entities, such as "chocolate bouquet Ferrero Rocher", present a challenge in discerning whether the user's search intent is focused on the chocolate, the bouquet or both. To enhance the intent understanding, existing works take attentions to extend the query information with the multi-granular sub-queries [8, 9] and user historical behavior sequence [8, 16]. However, they are still insufficient to understand the user's search intention. The sparse multi-granular sub-queries could bring a lot of noise for the long queries, such as the n-gram sub-query of the "chocolate bouquet Ferrero Rocher" can be "chocolate", but the user wants the bouquet with chocolate decorations rather than only chocolates. The user behaviors strongly depend on the product's side information within the sequences and designing highly complex sequences is challenging due to system performance constraints.

Figure 1: The visual imagination when a user searched for flowers on Meituan app.

In fact, the e-consumers typically form mental imagery of the product they tend to search for [2, 7], reflecting their specific purchase intentions such as shape, color, size and other details. Envisioning users' mental imagery of products can enhance the retrieval

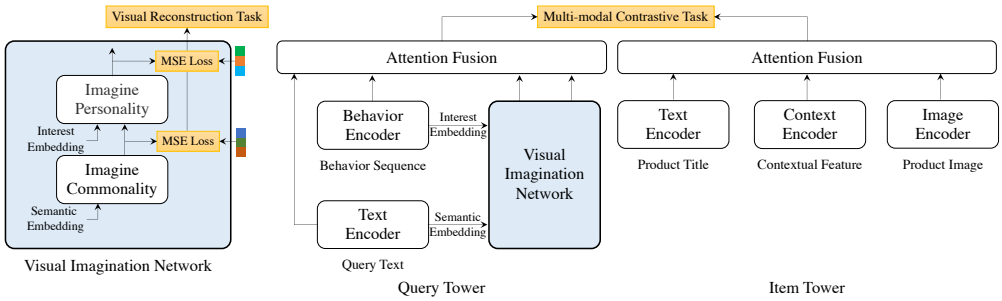

Figure 2: The overview of our proposed model VIER.

system's comprehension of in-depth search intents, both for short queries with vague intentions and long queries with semantic noise, thereby improving the final search results. Inspired by this, we present the **V**isual **I**magination **E**nhanced **R**etrieval (VIER) model to imagine the pseudo product images that the users prefer based on the search query and user behaviors. We introduce a visual imagination network designed to progressively construct both coarse-grained common and fine-grained personalized imaginations. As shown in Figure 1, by generating common mental imagery, the model can establish a universal baseline of understanding that is accessible to a diverse user searching the identical query. Building on this foundation, the generation of personalized mental imagery enables the model to adapt to individual user preferences, thereby facilitating a more tailored and nuanced understanding that aligns with each user's specific intent. To achieve this, we employ a multi-task learning approach that integrates a visual reconstruction task for generating product imagery with a multi-modal contrastive task aimed at refining final representations for enhanced retrieval performance.

In summary, the main contributions of this work are as follows:

- To the best of our knowledge, we are first to describe the mental product imagination based on the search query and user behaviors to enhance the search intention understanding in the EBR system.
- We propose a novel query tower with a lightweight visual imagination network and a multi-task learning strategy, designed to optimize product retrieval. Experimental offline results and online A/B tests both demonstrate the effectiveness of our model.

2 Method

In this section, we describe our proposed Visual Imagination Enhanced Retrieval (VIER) model as shown in Figure 2.

2.1 Query Tower

The query tower is used to encode the query and user information into a dense embedding vector, which is then used to retrieve the nearest product embeddings in the candidate embedding space.

2.1.1 Visual Imagination Network. Considering that e-consumers usually have a mental imagery of the product to be searched in their mind, we propose visual imagination network to explicitly construct the common and personalized imagination for the query extension. We first construct a common imagination based on the query semantic embedding $\boldsymbol{Q}_{text}$ using an Multilayer Perceptron (MLP) network, which captures the coarse-grained commonalities of the products related to the search query, as follows:

$$\boldsymbol{V}'_{cimg} = \mathrm{MLP_c}(\boldsymbol{Q}_{text}), \tag{1}$$

where we consider the mean of the image embeddings from all n products clicked in response to the given query to be the ground-truth representation of the common imagination, as follows:

$$\boldsymbol{V}_{cimg} = \frac{1}{n}\sum \boldsymbol{v}_i, \tag{2}$$

where each image embedding $\boldsymbol{v}_i$ is obtained by the frozen ResNet [4] rather than training end-to-end for computing efficiency.

Then we predict the fine-grained personalized imagination embedding $\boldsymbol{V}'_{pimg}$ based on the common imagination $\boldsymbol{V}'_{cimg}$ and the user's historical interest embedding $\boldsymbol{H}_{user}$ as follows:

$$\boldsymbol{V}'_{pimg} = \mathrm{MLP_p}(\boldsymbol{H}_{user}, \boldsymbol{V}'_{cimg}), \tag{3}$$

where $\boldsymbol{H}_{user}$ and $\boldsymbol{V}'_{cimg}$ are directly concatenated as the input, and $\mathrm{MLP_p}$ is another MLP network. We define the image embedding of the next product clicked by the user as the ground truth, denoted by $\boldsymbol{V}_{pimg}$.

The query semantic embedding $\boldsymbol{Q}_{text}$ and the user's interest embedding $\boldsymbol{H}_{user}$ are introduced in the following Section 2.1.2 and Section 2.1.3 respectively.

2.1.2 Text Encoder. We encode the textual query with a knowledge-enhanced Bidirectional Encoder Representation from Transformers (BERT) model similar to ERNIE [13]. Specifically, we combine the phrase-level entity embeddings with the character-level token embeddings to serve as the input in order to fully encode the entity knowledge of the query, and we take the output of the [CLS] token from the last layer as the query semantic embedding, as follows:

$$\boldsymbol{Q}_{text} = \mathrm{BERT_{wp}}(\boldsymbol{q}_{word}, \boldsymbol{q}_{phrase}), \tag{4}$$

where $\boldsymbol{q}_{word}$ denotes the character-level token embedding of the query. $\boldsymbol{q}_{phrase}$ denotes the embedding of recognized entities in the text, including brand, taste, color and core entities, which is pretrained by TransH method [15].

2.1.3 Behavior Encoder. Given a user history behavior sequence $\{c_1, \cdots, c_j, \cdots, c_m\}$ on the query, where c_j denotes one of the historical-clicked product id, we compute each embedding $\boldsymbol{h}_j$ via an MLP network as follows:

$$\boldsymbol{h}_j = \mathrm{MLP_b}(\mathrm{embed}(\mathrm{hash}(c_j))), \tag{5}$$

where embed$(\cdot)$ is the EmbeddingBag [12], hash$(\cdot)$ is the hash function (*e.g.* Murmur hash [14]). We acquire the interest embedding $\boldsymbol{H}_{user}$ with the sum pooling of the weighted $\boldsymbol{h}_j$ via self-attention mechanism as follows:

$$\boldsymbol{H}_{user} = \sum(\boldsymbol{h}_j \text{ attention}(\boldsymbol{h}_1, \cdots, \boldsymbol{h}_j, \cdots, \boldsymbol{h}_m)). \tag{6}$$

2.1.4 Multi-modal Attention Fusion. To effectively fuse the multi-modal embeddings of query, user history behaviors and imagined product image, we experiment with different fusion approaches including the concatenation fusion and self-attention fusion, which is generalized as the function Fusion$(\cdot)$ for simplicity:

$$\boldsymbol{Q} = \text{Fusion}_{\text{q}}(\boldsymbol{Q}_{text}, \boldsymbol{H}_{user}, \boldsymbol{V}'_{cimg}, \boldsymbol{V}'_{pimg}). \tag{7}$$

2.2 Item Tower

For the item tower, we encode the product representation with the multi-modal information including the title, image and contextual features (*e.g.* product id). The semantic embedding $\boldsymbol{P}_{text}$ and the image embedding $\boldsymbol{V}_{pimg}$ are encoded in the same way as the query tower. For the contextual embedding $\boldsymbol{H}_{context}$, we apply the EmbeddingBag [12] after hashing the product id and then feed it to context encoder. Finally, the same multi-modal fusion strategy is used to get the product representation as follows:

$$\boldsymbol{P} = \text{Fusion}_{\text{p}}(\boldsymbol{P}_{text}, \boldsymbol{H}_{context}, \boldsymbol{V}_{pimg}). \tag{8}$$

2.3 Training

We train our VIER model using the multi-task learning strategy, which encompasses a visual reconstruction task $\mathcal{L}_{vrt}$ for envisioning product imagery, and a multi-modal contrastive task $\mathcal{L}_{mct}$ for enhancing multi-modal retrieval:

$$\mathcal{L}(\theta) = \delta_1 \mathcal{L}_{vrt} + \delta_2 \mathcal{L}_{mct}, \tag{9}$$

where θ denotes all the learnable parameters of VIER, the task balance hyperparameters δ_1 and δ_2 are constrained such that $\delta_1 + \delta_2 = 1$.

2.3.1 Visual Reconstruction Task. We train the visual imagination network with the visual reconstruction loss $\mathcal{L}_{vrt}$, which is composed of a common imagination loss $\mathcal{L}_{ci}$ and a personalized imagination loss $\mathcal{L}_{pi}$ as follows:

$$\mathcal{L}_{vrt} = \lambda_1 \mathcal{L}_{ci} + \lambda_2 \mathcal{L}_{pi}, \tag{10}$$

where λ_1 and λ_2 are the loss hyperparameters. Specifically, we adopt the Mean Squared Error (MSE) training objective for both losses:

$$\mathcal{L}_{ci} = \text{MSE}(\boldsymbol{V}'_{cimg}, \boldsymbol{V}_{cimg}), \tag{11}$$

$$\mathcal{L}_{pi} = \text{MSE}(\boldsymbol{V}'_{pimg}, \boldsymbol{V}_{pimg}). \tag{12}$$

2.3.2 Multi-modal Contrastive Task. We conduct the contrastive learning [3] between the query and product based on the final multi-modal embeddings where the user-clicked <query, product> pairs constitute the positive samples, while the randomly sampled pairs within the mini-batch serve as the negative samples. Formally, we introduce the loss $\mathcal{L}_{mct}$ as follows:

$$\mathcal{L}_{mct} = -\frac{1}{N} \sum_{i=1}^{N} \log \frac{\exp(\cos(\boldsymbol{Q}_i, \boldsymbol{P}_i)/\tau)}{\sum_{j=1}^{N} \exp(\cos(\boldsymbol{Q}_i, \boldsymbol{P}_j)/\tau)}, \tag{13}$$

where N is the batch size and τ is the temperature parameter.

3 Experiments

3.1 Experimental Setup

3.1.1 Dataset. We randomly crawled 10 million user-clicked <query, product> pairs from the online Meituan search logs over a one-month period as the training dataset. For the evaluation, we sampled 1 million clicked pairs from the logs for the following week to ensure that the data was unseen during the training phase. Additionally, we collected 10k human-rated pairs comprising an equal number of relevant and irrelevant data to assess the relevance.

3.1.2 Offline Evaluation Metrics. We evaluate different models on the unseen evaluation datasets with Recall@K and ROC AUC of relevance as follows:

- **Recall@K**: To emulate the location-based service (LBS) scenario, we employ Faiss [5] to retrieve the top-K products within the 5×5 km^2 geohashed [11] locations using the query embedding and check whether the ground truth is among these top-K products. *Due to shallow browsing in LBS scenario, Recall@1 is our primary metric for rigorous offline evaluation, capturing both personalization and relevance.*
- **ROC AUC**: We calculate the ROC AUC score on the relevance evaluation dataset to assess the model's ability to recall semantically relevant products.

3.1.3 Baselines. We compare our model with both the text-only models and multi-modal models as follows:

- **BERT-base** [3]: A text-only 12-layer BERT model which only uses the original query and product title as the input and is trained with the contrastive learning task.
- **Que2Search** [9]: A multi-modal retrieval model which extends the query with the multi-granularity sub-queries and contextual features for the better retrieval results.
- **Que2Search+behavior**: Enhancing the query tower of the Que2Search model with the same behavior encoder as VIER.

3.1.4 Implementation Details. In the visual imagination network, the commonality encoder and the personality encoder are both 3-layer MLP networks. For the text encoder, we use a 12-layer BERT model with the output dimension of 768, which is initialized with the checkpoint pretrained on Meituan dataset. We use a 1-layer MLP as the context encoder. The image encoder is frozen in our model, which is the ResNet-50 [4] pretrained on the ImageNet [1] dataset. We utilize the historical-clicked products on the query within 3 months as the behavior sequence and segment the sequence using a window size of 50. The final multi-modal embeddings from query tower and item tower are mapped to 32-dimensional space using two 1-layer MLP networks for the online retrieval. For the task balance weights in Eq. 9, we determine the optimal weight through the grid search method. The loss weights in Eq. 10 are set to 1 for simplicity with a step size of 0.1. The temperature parameter τ in Eq. 13 is set as 0.05 in all experiments. We train the model with the Adam [6] optimizer with a learning rate of 2e-4 and a batch size of 64.

3.2 Offline Results

3.2.1 Main Result. As shown in Table 1, our proposed VIER model outperforms all baseline models in terms of Recall@1 and ROC AUC

Table 1: Comparison with the baseline methods on the evaluation dataset.

Methods	Recall@1	ROC AUC
BERT-base [3]	$0.7131^{(-2.66\%)}$	$0.8366^{(-3.16\%)}$
Que2Search [9]	$0.7242^{(-1.15\%)}$	$0.8500^{(-1.61\%)}$
Que2Search+behavior	$0.7259^{(-0.91\%)}$	$0.8483^{(-1.80\%)}$
VIER (ours)	**0.7326**	**0.8639**

on the evaluation dataset. Our model surpasses the best baseline with a relative boost of 0.91% in Recall@1 and 1.80% in ROC AUC, indicating its superior product retrieval capabilities.

To analyze the improvement of VIER on short and long queries, we show the Recall@1 relative improvements with the increasing length of query in Figure 3. Our results demonstrate that our model significantly improves Recall@1 for short queries (single-word queries) by 1.56% and for long queries (queries with six or more words) by 1.24%. This suggests that our model is especially effective in improving extreme scenarios where understanding the user's search intent is challenging.

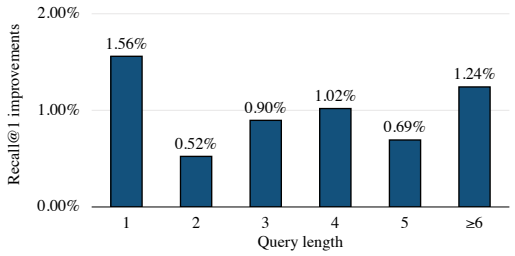

Figure 3: Relative improvements in Recall@1 of VIER compared to the best baseline as query length increases.

3.2.2 Ablation Study. To analyze the contributions of each component in our model, we conduct the ablation studies presented in Table 2. With the removal of the imagination network, our model experiences a 0.55% decrease in Recall@1, underscoring the significance of visual product imagination within the query tower. Both the commonality embedding and the personality embedding are essential, yielding improvements of 0.16% and 0.15%, respectively. Regarding the comparison of multi-modal fusion approaches, attention fusion outperforms simple concatenation fusion by 0.15%. Furthermore, within the text encoder of our model, the knowledge-enhanced BERT achieves a notable 0.61% improvement over the basic BERT which has the same architecture as BERT-base.

3.2.3 Loss Weights. We find that setting the weights for the visual reconstruction task (δ_1) and the multi-modal contrastive task (δ_2) to 0.4 and 0.6, respectively, yields the best result. This indicates that the visual reconstruction task indeed contributes positively to the overall performance. Additionally, when the weight δ_1 is too low, the model is unable to effectively learn the visual representation on the query side, thus failing to provide the intended enhancement for the query. Conversely, when the weight δ_1 is set too high, it leads to poor performance of the retrieval-focused contrastive learning task, resulting in sub-optimal retrieval outcomes.

Table 2: Ablation studies of each component in the VIER model.

Methods	Recall@1
w/o imagination network	$0.7286^{(-0.55\%)}$
w/o commonality embedding	$0.7314^{(-0.16\%)}$
w/o personality embedding	$0.7315^{(-0.15\%)}$
w/o attention fusion (concatenation fusion)	$0.7315^{(-0.15\%)}$
w/o knowledge-enhanced BERT (basic BERT)	$0.7281^{(-0.61\%)}$
VIER (our full model)	**0.7326**

3.2.4 Case Study. In the case study, it was discovered that when users searched for a short query such as "flower", VIER effectively learned the visual characteristics through the enhancement of visual imagination, thus identifying the users' visual preferences, such as flowers wrapped in black paper and adorned with a crown. In contrast, the best baseline did not capture these visual nuances, leading to predictions that failed to match the products that the user was most likely to select.

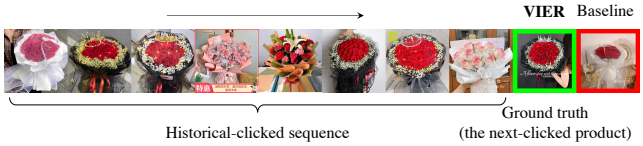

Figure 4: The product images of the historical-clicked sequence when searching for "flower", along with the models' predictions. The VIER model accurately recalled the ground truth at the top 1, while the best baseline model retrieved a product that the user didn't click on.

3.3 Online A/B Test

We conducted the online A/B test on Meituan sponsored-search system over a period of one month. The performances are evaluated by the revenue, clicks and click-through rate (CTR). It shows that VIER improves the 4.8% online revenue of sponsored-search system with a significant increase in clicks (+4.4%) and CTR (+2.3%) compared to the best baseline, which demonstrates that our model retrieves more high-quality products.

4 Conclusion

In this paper, we introduce a novel model, VIER, that enhances multi-modal retrieval with visual imagination to better understand user search intent. By simulating the mental imagery that e-consumers have of desired products, our model enriches query information to better reflect specific intentions. Our model notably boosts Recall@1 in hard search scenarios and increases system revenue by 4.8%, clicks by 4.4%, and CTR by 2.3% on Meituan sponsored-search system. This approach opens new research possibilities in EBR by incorporating mental imagery, and we hope our work will inspire further exploration in this area.

References

[1] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 248–255. https://doi.org/10.1109/CVPR.2009.5206848

[2] Jennifer Escalas. 2013. Imagine yourself in the product: Mental simulation, narrative transportation, and persuasion. *Journal of Advertising* 33 (03 2013), 37–48. https://doi.org/10.1080/00913367.2004.10639163

[3] Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. SimCSE: Simple Contrastive Learning of Sentence Embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 6894–6910. https://doi.org/10.18653/v1/2021.emnlp-main.552

[4] Kaiming He, X. Zhang, Shaoqing Ren, and Jian Sun. 2015. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015), 770–778.

[5] Jeff Johnson, Matthijs Douze, and Hervé Jégou. 2019. Billion-scale similarity search with GPUs. *IEEE Transactions on Big Data* 7, 3 (2019), 535–547.

[6] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun (Eds.). http://arxiv.org/abs/1412.6980

[7] Aurély Lao. 2013. Mental imagery and its determinants as factors of consumers emotional and behavioural responses: Situation analysis in online shopping. *Recherche et Applications en Marketing (English Edition)* 28, 3 (2013), 58–81. https://doi.org/10.1177/2051570713505479

[8] Sen Li, Fuyu Lv, Taiwei Jin, Guli Lin, Keping Yang, Xiaoyi Zeng, Xiao-Ming Wu, and Qianli Ma. 2021. Embedding-Based Product Retrieval in Taobao Search. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery amp; Data Mining* (Virtual Event, Singapore) *(KDD '21)*. Association for Computing Machinery, New York, NY, USA, 3181–3189. https://doi.org/10.1145/3447548.3467101

[9] Yiqun Liu, Kaushik Rangadurai, Yunzhong He, Siddarth Malreddy, Xunlong Gui, Xiaoyi Liu, and Fedor Borisyuk. 2021. Que2Search: Fast and Accurate Query and Document Understanding for Search at Facebook. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery amp; Data Mining* (Virtual Event, Singapore) *(KDD '21)*. Association for Computing Machinery, New York, NY, USA, 3376–3384. https://doi.org/10.1145/3447548.3467127

[10] Alessandro Magnani, Feng Liu, Suthee Chaidaroon, Sachin Yadav, Praveen Reddy Suram, Ajit Puthenputhussery, Sijie Chen, Min Xie, Anirudh Kashi, Tony Lee, and Ciya Liao. 2022. Semantic Retrieval at Walmart. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (Washington DC, USA) *(KDD '22)*. Association for Computing Machinery, New York, NY, USA, 3495–3503. https://doi.org/10.1145/3534678.3539164

[11] Gustavo Niemeyer. 2008. Geohash. https://web.archive.org/web/20080305223755/http://blog.labix.org/#post-85. Retrieved on 13 May 2023.

[12] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic Differentiation in PyTorch. In *NIPS 2017 Workshop on Autodiff* (Long Beach, California, USA). https://openreview.net/forum?id=BJJsrmfCZ

[13] Yu Sun, Shuohuan Wang, Yu-Kun Li, Shikun Feng, Xuyi Chen, Han Zhang, Xin Tian, Danxiang Zhu, Hao Tian, and Hua Wu. 2019. ERNIE: Enhanced Representation through Knowledge Integration. *CoRR* abs/1904.09223 (2019). arXiv:1904.09223 http://arxiv.org/abs/1904.09223

[14] Tanjent. 2008. MurmurHash first announcement. https://tanjent.livejournal.com/2008/03/03/. Retrieved on 13 May 2023.

[15] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge Graph Embedding by Translating on Hyperplanes. *Proceedings of the AAAI Conference on Artificial Intelligence* 28, 1 (Jun. 2014). https://doi.org/10.1609/aaai.v28i1.8870

[16] Han Zhang, Songlin Wang, Kang Zhang, Zhiling Tang, Yunjiang Jiang, Yun Xiao, Weipeng Yan, and Wen-Yun Yang. 2020. Towards Personalized and Semantic Retrieval: An End-to-End Solution for E-Commerce Search via Embedding Learning. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Virtual Event, China) *(SIGIR '20)*. Association for Computing Machinery, New York, NY, USA, 2407–2416. https://doi.org/10.1145/3397271.3401446

[17] Jianjin Zhang, Zheng Liu, Weihao Han, Shitao Xiao, Ruicheng Zheng, Yingxia Shao, Hao Sun, Hanqing Zhu, Premkumar Srinivasan, Weiwei Deng, Qi Zhang, and Xing Xie. 2022. Uni-Retriever: Towards Learning the Unified Embedding Based Retriever in Bing Sponsored Search. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (Washington DC, USA) *(KDD '22)*. Association for Computing Machinery, New York, NY, USA, 4493–4501. https://doi.org/10.1145/3534678.3539212