Co-visitation Meets Token Alignment for Next Product Title Generation

Chenwang Wu[1], Leyan Deng[1], Zhihao Zhu[1], Defu Lian[1]*
[1] University of Science and Technology of China
Hefei, Anhui, China
{wcw1996,dleyan,zzh98}@mail.ustc.edu.cn,liandefu@ustc.edu.cn

ABSTRACT

The title gives users a quick description of the product, which may help improve various downstream tasks of recommendations. Existing work utilizes language models with limited generation capabilities, which may lead to issues such as excessively brief and inefficient titles being produced. To this end, we proposed an efficient method to generate the next product title, consisting of the co-visitation recommendation module and the token alignment module. The co-visitation module gets the recommended title through a product co-visitation graph, while the token alignment module generates the final title by aligning the recommended title with the last interacted title to eliminate redundant tokens. Finally, we achieved high-quality titles and demonstrated competitive performance in the KDD CUP 2023 Tack3 competition (2nd place). Moreover, our method is very efficient and scalable, making it highly practical for large-scale systems.

CCS CONCEPTS

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

KEYWORDS

Session-based recommendation, title generation.

ACM Reference Format:
Chenwang Wu, Leyan Deng, Zhihao Zhu, Defu Lian. 2018. Co-visitation Meets Token Alignment for Next Product Title Generation. In *Proceedings of Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23).* ACM, New York, NY, USA, 4 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

In the digital era, recommendation technology plays a crucial role in e-commerce platforms [2]. It accurately understands user preferences by modeling their historical behavior. Product titles, which provide quick descriptions of product features and characteristics to users, can enhance various downstream tasks of recommendations, including cold-start recommendations [12] and navigation. Existing work [9, 15] utilizes language models with limited generation capabilities, which may lead to issues such as excessively brief and inefficient titles being produced.

The title, as a concise representation of a product, usually fulfills three requirements experientially: (1) captivating, (2) accurately depicts user preferences, and (3) accurately conveys key information. To address this, we design a simple yet efficient solution called "co-visitation meets token alignment" for generating the next product in recommendations. Specifically, it comprises two components: the co-visitation recommendation module and the token alignment module. First, the co-visitation module constructs a straightforward co-visitation graph, which intuitively captures correlations between products. Second, we align the recommended title based on the co-visitation module with recently interacted titles. This alignment operation eliminates redundancy in titles and conveys key information better. The generated title is derived from titles that the user has previously interacted with or a subset thereof, ensuring its appeal to users. Despite its simplicity, our approach yields high-quality titles, earning us the second-place position in the competition among 1990+ participants in KDD CUP 2023 Task 3. Additionally, we emphasize that our proposed solution is highly efficient and can be seamlessly scaled up to large-scale systems, making it highly practical.

2 RELATED WORK

Session-based recommendation (SBR) aims to model users' short-term preference within anonymous sessions to predict the next item, which can be viewed as a sub-field of sequential recommendation [14]. In recent years, SBR has played an important role in e-commerce systems and attracted increasing attention [16]. Recurrent neural network (RNN) [5, 6] and attention mechanism [4, 8, 13] are widely applied to SBR due to their expressive power in sequence data. For example, Hidasi et al. [6] first introduced RNN into recommender systems to achieve item-to-item recommendations and used a new ranking loss function to train the RNNs. Furthermore, Wang et al. [13] proposed an attention-based model to identify the most relevant item in a session to the next item. Since there exist some obvious graph structures in recommender systems, some works used Graph neural network (GNN) [3, 10, 11] to capture the dependencies between items.

Product titles provide users with quick descriptions of product functionalities and features, which may help improve downstream tasks. While some previous work [9, 15] has focused on it, the generated titles often fall short in terms of length and diversity due to the limitations in their generation capabilities. Recently, the development of large language models (LLMs) [1] has shown great potential for title generation. With LLMs having been trained on a wide range of topics and styles, their contextual understanding

∗ Corresponding Author.

KDD '23, August 6–10, 2023, Long Beach, CA, USA

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

and abilities in creativity and few-shot learning could provide a continuous source of technical support for product title generation.

3 METHEDOLOGY

Titles serve as the first point of interaction between products and potential users, significantly influencing their decision to click on a product or continue browsing. Therefore, titles must fulfill the following requirements: (1) be engaging, (2) effectively capture user preferences, and (3) accurately convey key information.

To meet these requirements, we design two modules as follows:

- **Co-visitation Module**: Firstly, generating new titles directly may not guarantee that they will be captivating. To address this, we carefully select the generated titles from existing titles or their subsets. Secondly, in line with the requirement of depicting user preferences, we aim to develop a high-performing recommendation system. Taking these considerations into account, we propose a co-visitation module that recommends a suitable title as the candidate title for the final title generation.
- **Token Alignment Module**: The recommended titles may contain redundant information which does not fulfill the third condition. Furthermore, the evaluation metric BLEU has limitations in considering higher-level translation accuracy, such as context, grammar, and semantics. Even similar titles can potentially receive low scores. Therefore, we align the recommended titles with the last interacted title by taking the intersection of tokens. This allows us to retain more crucial information.

3.1 Co-visitation Module

A well-performing recommender system is a crucial prerequisite for generating high-quality product titles. This can aid the system in gaining a deeper understanding of user preferences, enabling the formulation of various personalized titles. In our work, we choose a straightforward and intuitive recommendation strategy known as covisitation recommendation based on the BLEU metric.

BLEU solely evaluates the consistency with reference titles and fails to capture human subjective judgment. This means that even if the recommended product is highly similar to the ground truth, factors such as different token order or synonyms can lower the BLEU score. For example, the BLEU score for "apple" and "apples" is 0. We cannot tolerate such risks. Therefore, we aim for the recommendation model to distinguish different products' recommended confidence effectively. If the confidence of a product is significantly higher than others, we consider it highly likely to be correct and use its title as a candidate for the final title generation. Otherwise, even if it is correct, we discard it due to the difficulty in bearing such risks. Regarding classical session-based recommender systems like SASRec [8] and GRU4Rec [6], although they achieve satisfactory performance, we found that they tend to be overly confident. That is, the confidence of the recommended product (often exceeding 0.99) is consistently much higher than other values, regardless of the recommendation accuracy. In addition, these models often act as black boxes that are difficult for humans to understand, further reducing our trust in the recommendation results.

Therefore, we aim to utilize a simple, transparent model with well-distinguishable recommendation scores, making co-visitation recommendation our choice. Co-visitation is defined as the occurrence of two events being clicked by the same user within a certain time interval, and its idea is intuitive. For example, if a user tends to click on iPhone 14, they would likely be interested in iPhone 14 Pro Max. This would result in significantly higher connection weights for iPhone 14 and iPhone 14 Pro Max than other products. In this scenario, when a new user clicks on iPhone 14, it becomes natural to recommend iPhone 14 Pro Max to him.

Constructing co-visitation is very critical. As we aim to predict the final product, the earlier clicked products within the session may not provide positive assistance due to timeliness issues. Therefore, when constructing the co-visitation graph, we only extract each session's most recent interacted records. Formally, for each session $s \in \mathcal{D}$, we only consider the co-visitation graph $\mathcal{A}$ based on its last k interaction records $last(s, k)$, we set k to 5 in our work. Additionally, co-visitation weights should be higher for products that are closer in relative position. Taking these considerations into account, we formalize the co-visitation graph $\mathcal{A}$'s construction rules for the weight between product p and q are as follows:

$$\mathcal{A}_{p,q} = \sum_{s \in \mathcal{D}} \sum_{i=1}^{min(5,|s|)} \sum_{j=1}^{min(5,|s|)} \sum_{s_i = p, s_j = q} 1.1 - |i - j| \cdot 0.2. \quad (1)$$

During the inference stage, for a test session s, we select the last item s_{-1} in it and obtain all products $A_{s_{-1}}$ co-visiting to it. As we discussed earlier, we only consider the recommended product if its recommendation score is exceptionally high. We use the 6-sigma rule to determine the threshold h_s:

$$h_s = mean(\mathcal{A}_{s_{-1}}) + 6 \cdot std(\mathcal{A}_{s_{-1}}). \quad (2)$$

If the highest weight $max(\mathcal{A}_{s_{-1}}) > h_s$, we consider the title corresponding to that product as a recommended title $\mathcal{T}_{rec}$, that is, $\mathcal{T}_{rec}(s) = \mathcal{T}(\arg\max(\mathcal{A}_{s_{-1}})$, to generate the final title $\mathcal{T}_{pred}(s)$ (will be introduced in Section 3.2), where $\mathcal{T}(i)$ denotes the title of product i. However, if $max(\mathcal{A}_{s_{-1}}) \leq h_s$, to mitigate the risk of prediction errors, we directly select the last interacted title as the final prediction $\mathcal{T}_{pred}(s) = \mathcal{T}(s_{-1})$.

3.2 Token Alignment

Although we prioritize titles with high confidence to mitigate risks, they often contain redundant tokens (such as size, color, etc.), making it challenging to convey key information accurately. Therefore, we aim to extract the key information from the titles obtained from the co-visitation module. We find that using the title of the last interacted product as the predicted title can also yield satisfactory results, and the last interacted titles and the recommended title by the co-visitation module are often highly similar. For example, if the last interacted title is "Men's 3/4 Length Jersey Training Active 3/4 Woven Pants, Blue," the recommended title might be "Men's 3/4 Length Jersey Training Active 3/4 Woven Pants, Black." Hence, by taking the intersection of the last interacted title and the recommended title, we can remove non-essential tokens, that is, color in this case. To achieve this, we propose the token alignment module, which involves taking the intersection of the last interacted title $\mathcal{T}(s_{-1})$ and the recommended title $\mathcal{T}_{rec}(s)$, denoted as

$$T_{align}(s) = join(tokenize(\mathcal{T}(s_{-1})) \cap tokenize(\mathcal{T}_{rec}(s))), \quad (3)$$

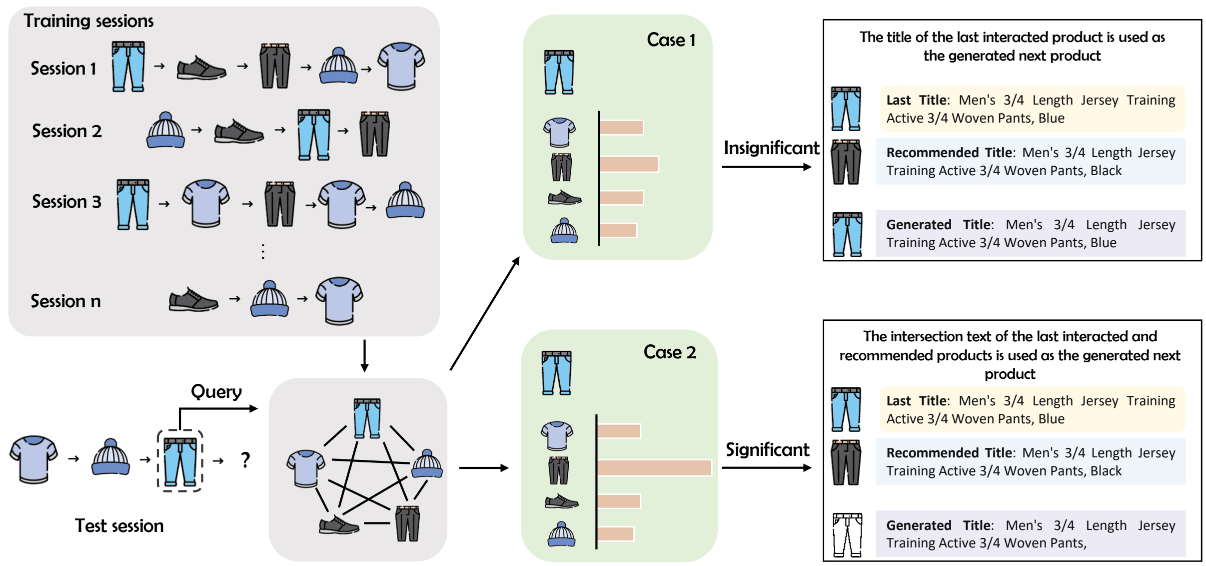

Figure 1: The framework of our method.

where $tokenize(\cdot)$ is tokenized function, and $join(\cdot)$ aims to combine all tokens into one sentence.

Indeed, there may be cases where the extracted tokens after alignment are too few, resulting in shorter titles that are penalized with lower BLEU scores. To address this, we introduce a threshold. If the number of tokens in the aligned title $\mathcal{T}_{align}(s)$ exceeds this threshold, we consider the aligned title as the final predicted title $\mathcal{T}_{pred}(s)$. Otherwise, we still select the last interacted title as the final prediction. Here, we set the threshold as 50% of the tokens in the recommended title. That is,

$$\begin{cases} \mathcal{T}_{pred}(s) = \mathcal{T}_{align}(s), \quad if \ \ |\mathcal{T}_{align}(s)| > 0.5 \cdot |\mathcal{T}_{rec}(s)|; \\ \mathcal{T}_{pred}(s) = \mathcal{T}(s_{-1}), \quad otherwise. \end{cases} \tag{4}$$

By introducing this threshold, we ensure that predicted titles containing an adequate number of tokens are not penalized, preventing a significantly low BLEU score. It also enables the titles to convey sufficient critical information. The framework and specific algorithm process can be seen in Fig. 1 and Alg. 1.

4 EXPERIMENTS

4.1 Datasets

We use the Multilingual Shopping Session Dataset [7], which collected customer sessions from six different language environments, including English, German, Japanese, French, Italian, and Spanish. It consists of two primary components: user sessions and product attributes. The user sessions are sequences of products with which users interacted, organized chronologically. The statistics information is shown in Table 1. The product attributes encompass various details such as product titles, local currency prices, brands, colors, and descriptions.

Algorithm 1: The complete process of our algorithm.

```
Construct the co-visitation graph based on Eq. 1;
for each test session s do
    Calculate the confident threshold h_s based on Eq. 2;
    if max(A_{s_{-1}}) > h_s then
        Get the aligned title T_align(s) based on Eq. 3;
        if |T_align(s)| > 0.5 · |T_rec(s)| then
            T_pred(s) = T_align(s);
        end
        else
            T_pred(s) = T(s_{-1});
        end
    end
    else
        T_pred(s) = T(s_{-1});
    end
end
return T_pred
```

Table 1: Statistics for the datasets

Language (Locale)	# Sessions	# Products (ASINs)
German (DE)	1111416	513811
Japanese (JP)	979119	389888
English (UK)	1182181	494409
Spanish (ES)	89047	41341
French (FR)	117561	43033
Italian (IT)	126925	48788

4.2 Case Study

Due to the unavailability of ground truth, we cannot provide a detailed experimental analysis. Therefore, we only present a few examples of the generated titles, as shown in Table 2. As mentioned

Table 2: Examples of generated titles.

Locale	Last Interacted Title	Recommended Title	Predicted Title
DE	Dampflion Aromakonzentrat Checkmate - Black Bishop, zum Mischen mit Basisliquid für e-Liquid, 0.0 mg Nikotin, 10 ml	Dampflion Aromakonzentrat Checkmate - White Bishop, zum Mischen mit Basisliquid für e-Liquid, 0.0 mg Nikotin, 10 ml	Dampflion Aromakonzentrat Checkmate - Bishop, zum Mischen mit Basisliquid für e-Liquid, 0.0 mg Nikotin, 10 ml
JP	Amazonベーシック ノート クラシック ノートブック Lサイズ 無地	Amazonベーシック ノート クラシック ノートブック Lサイズ 横罫	Amazonベーシック ノート クラシック ノートブック Lサイズ
UK	Bosch Home and Garden Cordless Combi Drill UniversalImpact 18 (1 battery, 18 Volt System, in carrying case)	Bosch Home and Garden Cordless Combi Drill UniversalImpact 18 (2 batteries, 18 Volt System, in carrying case)	Bosch Home and Garden Cordless Combi Drill UniversalImpact 18 (, 18 Volt System, in carrying case)
ES	Amazon Basics - Pizarra blanca magnética con bandeja para rotuladores y marco de aluminio, 60 cm x 45 cm	Amazon Basics - Pizarra blanca magnética con bandeja para rotuladores y marco de aluminio, 120 cm x 90 cm	Amazon Basics - Pizarra blanca magnética con bandeja para rotuladores y marco de aluminio, cm x cm
FR	JETech Coque pour iPhone 12/12 Pro 6,1 Pouces, Anti-Jaunissement étui de Protection Transparente Antichoc, Housse Case Cover Anti-Rayures (HD Clair)	JETech Coque Compatible avec iPhone 12/12 Pro 6,1 Pouces, étui de Protection Transparente Antichoc, Housse Case Cover Anti-Rayures (Bleu)	JETech Coque iPhone 12/12 Pro 6,1 Pouces, étui de Protection Transparente Antichoc, Housse Case Cover Anti-Rayures ()
IT	Sony Mdr-Ex15Ap - Cuffie In-Ear con Microfono, Auricolari in Silicone, Nero	Sony Mdr-Ex15Ap - Cuffie In-Ear con Microfono, Auricolari in Silicone, Blu	Sony Mdr-Ex15Ap - Cuffie In-Ear con Microfono, Auricolari in Silicone,

Table 3: Running Times (s)

Locale	DE	JP	UK	ES	FR	IT
Train	136.9					
Test	0.00007	0.00006	0.00008	0.00009	0.0001	0.00011

in Section 3.2, the recommended and last interacted titles are highly similar, which confirms the good performance of using the last interacted title as the predicted title to some extent (online BLEU score: 0.26553). Furthermore, we can intuitively feel that the predicted titles obtained through token alignment effectively preserve crucial information and minimize the decrease in BLEU scores caused by irrelevant information. Despite the seemingly simple combination of co-visitation recommendation and token alignment, this approach was carefully designed. The online evaluation performance is satisfactory, achieving the second-best score (BLEU: 0.27131) among all participating teams.

4.3 Efficiency Evaluation

In addition, we have provided the runtime of our model to evaluate its efficiency, as shown in Table 3. The "Train" column represents the time taken to construct the co-visitation graph, while the "Test" column indicates the average time of title generation per session. Clearly, our algorithm is highly efficient, highlighting its flexibility to scale up to large-scale systems and its practicality.

5 CONCLUSION

In this paper, we present an efficient solution called "co-visitation meets token alignment" for generating the next product title in recommender systems. The co-visitation module constructs a co-visitation graph to capture user preferences, while the token alignment module aligns the recommended title with recently interacted titles to further extract the key information. The evaluation results demonstrate its effectiveness, which resulted in high-quality titles and secured the second-place position in KDD CUP 2023 Task 3. Furthermore, the proposed solution is highly efficient and can be seamlessly scaled up to large-scale systems, making it practical for implementation in real-world e-commerce platforms.

REFERENCES

[1] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.

[2] Jin Chen, Ju Xu, Gangwei Jiang, Tiezheng Ge, Zhiqiang Zhang, Defu Lian, and Kai Zheng. 2021. Automated Creative Optimization for E-Commerce Advertising. In *Proceedings of the Web Conference 2021*. 2304–2313.

[3] Tianwen Chen and Raymond Chi-Wing Wong. 2020. Handling information loss of graph neural networks for session-based recommendation. In *Proceedings of KDD'20*. 1172–1180.

[4] Wanyu Chen, Fei Cai, Honghui Chen, and Maarten de Rijke. 2019. A dynamic co-attention network for session-based recommendation. In *Proceedings of CIKM'19*. 1461–1470.

[5] Balázs Hidasi and Alexandros Karatzoglou. 2018. Recurrent neural networks with top-k gains for session-based recommendations. In *Proceedings of CIKM'18*. 843–852.

[6] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2015. Session-based recommendations with recurrent neural networks. *arXiv preprint arXiv:1511.06939* (2015).

[7] Wei Jin, Haitao Mao, Zheng Li, Haoming Jiang, Chen Luo, Hongzhi Wen, Haoyu Han, Hanqing Lu, Zhengyang Wang, Ruirui Li, et al. 2023. Amazon-M2: A Multilingual Multi-locale Shopping Session Dataset for Recommendation and Text Generation. *arXiv preprint arXiv:2307.09688* (2023).

[8] Wang-Cheng Kang and Julian McAuley. 2018. Self-attentive sequential recommendation. In *ICDM'18*. IEEE, 197–206.

[9] Mansi Ranjit Mane, Shashank Kedia, Aditya Mantha, Stephen Guo, and Kannan Achan. 2020. Product title generation for conversational systems using BERT. *arXiv preprint arXiv:2007.11768* (2020).

[10] Zhiqiang Pan, Fei Cai, Wanyu Chen, Honghui Chen, and Maarten De Rijke. 2020. Star graph neural networks for session-based recommendation. In *Proceedings of CIKM'20*. 1195–1204.

[11] Ruihong Qiu, Jingjing Li, Zi Huang, and Hongzhi Yin. 2019. Rethinking the item order in session-based recommendation with graph neural networks. In *Proceedings of CIKM'19*. 579–588.

[12] Li Wang, Binbin Jin, Zhenya Huang, Hongke Zhao, Defu Lian, Qi Liu, and Enhong Chen. 2021. Preference-Adaptive Meta-Learning for Cold-Start Recommendation.. In *IJCAI*. 1607–1614.

[13] Shoujin Wang, Liang Hu, Longbing Cao, Xiaoshui Huang, Defu Lian, and Wei Liu. 2018. Attention-based transactional context embedding for next-item recommendation. In *Proceedings of AAAI'18*, Vol. 32.

[14] Shiwen Wu, Fei Sun, Wentao Zhang, Xu Xie, and Bin Cui. 2022. Graph neural networks in recommender systems: a survey. *Comput. Surveys* 55, 5 (2022), 1–37.

[15] Jian-Guo Zhang, Pengcheng Zou, Zhao Li, Yao Wan, Xiuming Pan, Yu Gong, and Philip S Yu. 2019. Multi-modal generative adversarial network for short product title generation in mobile e-commerce. *arXiv preprint arXiv:1904.01735* (2019).

[16] Xiaokun Zhang, Bo Xu, Liang Yang, Chenliang Li, Fenglong Ma, Haifeng Liu, and Hongfei Lin. 2022. Price does matter! modeling price and interest preferences in session-based recommendation. In *Proceedings of SIGIR'22*. 1684–1693.