Title Generation For Multilingual Session-Based Recommendation Systems: Project Proposal

Minh Duc Nguyen
Department of
Computer Science and Engineering
UNIST
ducnm@unist.ac.kr

Thu Phuong Nguyen
Department of
Computer Science and Engineering
UNIST
phuongnt@unist.ac.kr

Abstract

E-commerce is becoming more popular thanks to its convenience and its advanced recommendation systems. Despite that, there are few studies on session-based recommendation systems in the multilingual setting. In particular, engaging product title generation can help companies to personalize recommendations and advertisements. In this work, we introduce an overview of the problem and propose a comprehensive development plan for a title-generation framework utilizing Natural Language Processing techniques together with our current research progress.

1 Introduction

In recent years, e-commerce has gradually expanded its popularity and slowly taken over traditional shopping as the preferred method of buying goods because of its inherent convenience in this digital era. However, its success is not just due to its convenience, with more and more e-commerce platforms entering the market, they are constantly looking for ways to enhance users' shopping experiences. One of the most vital components when it comes to improving users' experience is the recommendation system as it is the fundamental functionality that makes such platforms convenient. Although there are numerous studies on state-of-the-art recommendation systems, few studies have investigated session-based recommendations, that is, given user previous interaction, recommend the next engaging product, in practical settings with imbalanced and multilingual data scenarios. In particular, predicting or generating product titles is extremely helpful when it comes to personalized recommendation systems or personalized advertisements, which will benefit e-commerce businesses by improving customer experience and increasing sales.

In this paper, we propose our tentative plan for developing a title-generation framework using natural processing techniques for better session-based recommendation systems. The system will receive users' previous interactions in a session and produce the title for the next engaging product. To successfully achieve the goal, we will utilize Amazon's "Multilingual Shopping Session Dataset" [1] consisting of millions of user sessions from six different locales: English, German, Japanese, French, Italian, and Spanish. With the use of natural language processing techniques and this comprehensive dataset, we aim to develop a highly accurate and efficient model.

The contributions of our work will be as follows:

- An exploratory data analysis on the real world "Multilingual Shopping Session Dataset".
- A deep learning framework using Large Language Models (LLM) to generate engaging product titles.
- A comprehensive empirical study on the efficiency of different LLMs.

Preprint. Under review.

2 Related Works

2.1 Literature Survey

The early appearances of deep learning in the research of recommendation systems suggest the use of Multi-Layer Perceptron (MLP) and Recurrent Neural Networks (RNN). He et al. [8] proposed Neural Collaborative Filtering to generate embeddings for each user and item. Hidasi et al. [9] proposed the use of Gated Recurrent Units [6] to capture the sequential properties of the session. Recently, most deep-learning approaches are heavily influenced by Transformer-based architecture [19] thanks to its self-attention innovation. Kang and McAuley [11] stacked this powerful architecture block to auto-regressively predict the next engaged item given N-length sequence.

These techniques could be helpful for title generation since we need information from previously interacted products. However, none of them could handle textual inputs, therefore, it would be hard to process multilingual information from the dataset if we want to directly utilize those. To address the multilingual problem, Long Short Term Memory was introduced [10]. Artetxe and Schwenk [2] proposed LASER, a toolkit that can generate cross-lingual sentence embeddings for over 90 languages by using a Bi-directional LSTM encoder and a traditional LSTM decoder. Mathur et al. [14] suggest using sequence-to-sequence models like Bi-directional RNN for title generation with transfer learning to address the data imbalance problem.

In recent approaches related to title generation for e-commerce platforms, Large Language Models (LLMs) are frequently used because of their state-of-the-art performance on various benchmarks in Natural Language Processing (NLP). Liu et al. [13] utilized GRU to produce embeddings and keywords generated from Transformer to provide recommendations given previous product titles. The main architecture of their keyword generation model is OpenNMT [12], a machine translation model capable of handling multiple languages. The model was trained on one day's worth of data from a real-world e-commerce platform with a similar number of entries to our data. Their proposed solution achieved state-of-the-art accuracy on top-MRR and top-Recall metrics compared to other baseline models. Bai et al. [3] was the first to use a generative approach for identifying product attributes. They concatenate the product title, the Google Product Taxonomy category, and the ground truth attributes into two sentences with special tokens to separate the different parts, then a base BERT [7] model was fine-tuned with a sequence-to-sequence objective to generate product names from the textual content of a product. Their model achieved the highest accuracy among the baseline models.

Although the above previous work has achieved great results, no discovery is found when it comes to efficiently generating engaging product titles, and there are only a few studies on how modern LLMs can contribute to this problem.

2.2 Pre-trained Large Language Models

In the past five years, the boom of deep learning in Natural Language Processing and the birth of the Transformers [19] architecture dictates the trend in how language models are designed. Bigger and deeper models are preferred since they are capable of producing human-like text and achieving state-of-the-art performance on most of the benchmarks. In 2018, Google and OpenAI research groups released the first generation of Large Language Models, namely, BERT [7] and GPT [16]. Following their work, LLMs are improved with deeper architectures and bigger datasets that contain multiple languages. Shortly after the release of BERT, its multilingual variation was introduced by the same group. Brown et al. [5] proposed GPT-3, the most capable LLM at the time, however, it was closed-source and people have to pay for API access. Shliazhko et al. [17] reproduced the GPT-3 architecture using GPT-2 sources and introduced mGPT with multilingual capability. BigScience Workshop [4] introduced BLOOM, an open-source LLM with similar sizes and performance to GPT-3. Touvron et al. [18] released a new LLM namely LLaMA with similar performance to GPT-3 but ten times smaller and faster.

We will conduct experiments with some of those LLMs to investigate their performance on this particular problem.

3 Dataset and Evaluation

3.1 Dataset

We will conduct an experiment on Amazon's "Multilingual Shopping Session Dataset" [1], which contains anonymized customer sessions and product attributes from six different locales: English, German, Japanese, French, Italian, and Spanish. The dataset statistics, including the total number of sessions and the number of products for each locale, are summarized in Figure 1.

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

Figure 1: Dataset statistics

There are about 1.55 million products in the dataset collected in multilingual languages with information about their locale, id, title, price, brand, color, size, model, material, author, and description, as can be seen in Figure 2. Most of the components in product attributes are unstructured, and even the same product ID, which is a unique Amazon Standard Identification Number (ASIN), could have different prices by locale. In addition, all the texture data has a varied range of options to represent because of the usage of space, bracket, and capitalized words making it hard to tokenize and model texture information.

	id	locale	title	price	brand	color	size	model	material	author	desc
0	B005ZSSN10	DE	RED DRAGON Amberjack 3 - Steel Tip 22 Gramm Wo...	30.95	RED DRAGON	NaN	NaN	RDD0089	NaN	NaN	Amberjacks Steel Dartpfeile sind verfügbar in ...
1	B08PRYN6LD	DE	Simply Keto Lower Carb* Schokodrops ohne Zucke...	17.90	Simply Keto	NaN	750 g (1er Pack)	NaN	NaN	NaN	NATÜRLICHE SÜSSE DURCH ERYTHRIT - Wir stelle...
2	B09MBZJ48V	DE	Sennheiser 508377 PC 5.2 Chat, Stilvolles Mult...	68.89	Sennheiser	Multi-Colour	One size	508377	Kunstleder	NaN	3.5 MM BUCHSE - Kann problemlos an Geräte mit ...
3	B08ZN6F26S	DE	AmyBenton Auto ab 1 2 3 ahre - Baby Aufziehbar...	18.99	Amy & Benton	Animal Car	NaN	2008B	aufziehauto 1 jahr	NaN	【Auto aufziehbar】: Drücken Sie einfach leicht ...
4	B094DGRV7D	DE	PLAYMOBIL - 70522 - Cavaliere mit grauem Pony	7.17	PLAYMOBIL	Nicht Zutreffend.	OneSize	70522	Polypropylen	NaN	Inhalt: 1 Stück

Figure 2: Product attributes in the dataset.

All user sessions are stored in a list of products that a user has engaged with in chronological order. Please refer to Figure 3 for more details.

	prev_items	next_item	locale
0	[B09W9FND7K, B09JSPLN1M]	B09M7GY217	DE
1	[B076THCGSG, B007MO8IME, B08MF65MLV, B001B4TKA0]	B001B4THSA	DE
2	[B0B1LGXWDS, B00AZYORS2, B0B1LGXWDS, B00AZYORS...	B0767DTG2Q	DE
3	[B09XMTWDVT, B0B4MZZ8MB, B0B7HZ2GWX, B09XMTWDV...	B0B4R9NN4B	DE
4	[B09Y5CSL3T, B09Y5DPTXN, B09FKD61R8]	B0BGVBKWGZ	DE

Figure 3: User sessions in the dataset

3.2 Evaluation Metrics

The dataset also contains human reference annotations for each entry in the testing data. Thus, we will proceed to use the standard evaluation metrics to evaluate the quality of the title generation, namely, Bilingual Evaluation Understudy (BLEU) [15]. Formally, BLEU could be mathematically represented as

$$\mathrm{BLEU} = \mathrm{BP} \times \exp\left(\sum_{n=1}^{N} w_n \log p_n\right), \tag{1}$$

where BP, N, w_n, and p_n is the brevity penalty, maximum n-gram length, weight assigned to each n-gram, and p_n is the precision score of each n-gram respectively. In particular, we will use $N = 4$, namely, BLEU-4, and a uniform distribution of weights $w_n = \dfrac{1}{N}$ for evaluation.

4 Tentative Project Plan

In this section, we put forward a preliminary and all-encompassing schedule for our project. This includes fundamental tasks, the timeline for each task, and the expected timeframe for completion of each phase. Please refer to table 1 to see the full picture of our strategy. Note that the table also includes those that have already been done by us prior to this proposal.

Order	Task	Schedule	Duration
1	Data EDA and pre-processing	3$^{\text{rd}}$ to 4$^{\text{th}}$ week	2 weeks
2	Literature Survey	5$^{\text{th}}$ to 6$^{\text{th}}$ week	2 weeks
3	Implementing baseline models & comparing efficiency	7$^{\text{th}}$ to 10$^{\text{th}}$ week	4 weeks
5	Finding the state of the art model	11$^{\text{th}}$ week	1 week
6	Improving the performance of the chosen model	12$^{\text{th}}$ to 13$^{\text{th}}$ week	2 weeks
7	Making final presentation	14$^{\text{th}}$ week	1 week
8	Finalize project report and & submission	15$^{\text{th}}$ week	1 week

Table 1: Project Plan

References

[1] Amazon. Multilingual Shopping Session Dataset. https://www.aicrowd.com/challenges/amazon-kdd-cup-23-multilingual-recommendation-challenge, 2023.

[2] Mikel Artetxe and Holger Schwenk. Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. *Transactions of the Association for Computational Linguistics*, 7:597–610, 2018.

[3] Xiao Bai, Lei Duan, Richard Wing Cheong Tang, Gaurav Batra, and Ritesh Agrawal. Improving text-based similar product recommendation for dynamic product advertising at yahoo. *Pro-*

ceedings of the 31st ACM International Conference on Information & Knowledge Management, 2022.

[4] BigScience Workshop. BLOOM. https://huggingface.co/bigscience/bloom, 2022.

[5] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, T. J. Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeff Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. *ArXiv*, abs/2005.14165, 2020.

[6] Kyunghyun Cho, Bart van Merrienboer, Çaglar Gülçehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In *Conference on Empirical Methods in Natural Language Processing*, 2014.

[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *ArXiv*, abs/1810.04805, 2019.

[8] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. *Proceedings of the 26th International Conference on World Wide Web*, 2017.

[9] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. Session-based recommendations with recurrent neural networks. *CoRR*, abs/1511.06939, 2015.

[10] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9: 1735–1780, 1997.

[11] Wang-Cheng Kang and Julian McAuley. Self-attentive sequential recommendation. *2018 IEEE International Conference on Data Mining (ICDM)*, pages 197–206, 2018.

[12] Guillaume Klein, François Hernandez, Vincent Nguyen, and Jean Senellart. The OpenNMT neural machine translation toolkit: 2020 edition. In *Proceedings of the 14th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Track)*, pages 102–109, Virtual, October 2020. Association for Machine Translation in the Americas. URL https://aclanthology.org/2020.amta-research.9.

[13] Yuanxing Liu, Zhaochun Ren, Weinan Zhang, Wanxiang Che, Ting Liu, and Dawei Yin. Keywords generation improves e-commerce session-based recommendation. *Proceedings of The Web Conference 2020*, 2020.

[14] Prashant Mathur, Nicola Ueffing, and Gregor Leusch. Multi-lingual neural title generation for e-commerce browse pages. *ArXiv*, abs/1804.01041, 2018.

[15] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Annual Meeting of the Association for Computational Linguistics*, 2002.

[16] Alec Radford and Karthik Narasimhan. Improving language understanding by generative pre-training. 2018.

[17] Oleh Shliazhko, Alena Fenogenova, Maria Tikhonova, Vladislav Mikhailov, Anastasia Kozlova, and Tatiana Shavrina. mgpt: Few-shot learners go multilingual. *ArXiv*, abs/2204.07580, 2022.

[18] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aur'elien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. Llama: Open and efficient foundation language models. *ArXiv*, abs/2302.13971, 2023.

[19] Ashish Vaswani, Noam M. Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *ArXiv*, abs/1706.03762, 2017.