Scaling Use-case Based Shopping using LLMs

Sachin Farfade
sfarfade@amazon.com
Amazon, India

Sachin Vernekar
svernek@amazon.com
Amazon, India

Vineet Chaoji
vchaoji@amazon.com
Amazon, India

Rajdeep Mukherjee
rajdeep1989.iitkgp@gmail.com
IIT Kharagpur, India

ABSTRACT

Products on e-commerce websites are usually organized based on seller-provided product attributes. Customers looking for a product typically have certain needs or product use-cases in mind, for e.g., a headphone for gym classes, or a printer for a small business. However, they often struggle to map these use-cases to product attributes and subsequently fail to find the product they need. In this talk, we present a use-case based shopping (UBS) ML system that facilitates use-case based customer experiences (CXs). The UBS system recommends dominant product use-cases to customers along with most relevant products for those use-cases. Use-cases and their definitions vary across product categories and marketplaces (MPs). This makes training supervised models for thousands of e-commerce categories and multiple MPs infeasible by collecting large amount training data needed to train these models. In this talk, we present our work on scaling the UBS model by instruction tuning an LLM for our task.

CCS CONCEPTS

• **Applied computing** → **Online shopping**; • **Computing methodologies** → **Large Language Models**; *Natural language generation.*

KEYWORDS

e-commerce, use-case extraction, instruction tuning, large language models, generalizability, task adaptation, Claude

ACM Reference Format:
Sachin Farfade, Sachin Vernekar, Vineet Chaoji, and Rajdeep Mukherjee. 2024. Scaling Use-case Based Shopping using LLMs. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining (WSDM '24), March 4–8, 2024, Merida, Mexico.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3616855.3635748

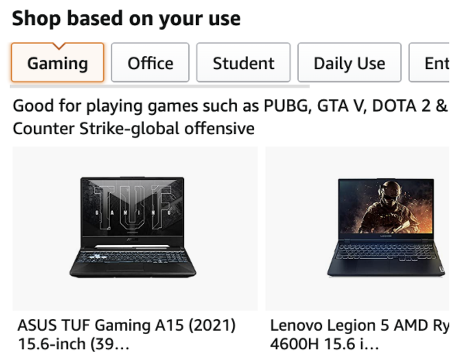

Figure 1: Example CX based on UBS for *laptop* category

1 INTRODUCTION

Customers are usually not aware of the exact product to buy and they typically start with a product category and a use-case (e.g., "headphones for *running*", "laptops for *online classes*" etc.) they need the product for. A customer wanting a printer for their small/medium business does not necessarily understand whether to get a laser-jet printer (with a throughput of 30 pages per second), which results in cost savings. In the physical store, the storekeeper engages with the customers in a conversation, understands their use-case and recommends an appropriate product for them. This assistance is essential in helping customers make confident purchase decisions. While e-commerce websites have a large selection of products to cater to diverse customer needs, they struggle to find the right products for their use-case in absence of such an assistance. In this talk, we present UBS, a system that assists customers shop based on their usage with use-case based CXs (Figure 1). Scaling the UBS system to e-commerce scale is challenging, as it is expensive and time consuming to collect training data needed to train the models. Recently, instruction tuned generative LLMs [6] have been shown to perform well on unseen tasks [1, 2, 4], as well as open domain conversations [3, 5]. In this talk, we present our key learnings on scaling the UBS system using limited training data with LLMs.

2 UBS SYSTEM

Figure 2 shows the ML pipeline for UBS. It consists of three main components: 1) Use-case phrase identification (PI): Given a product category, this module identifies relevant use-case phrases from customer reviews, Q&As, product catalogue and search queries, 2) Dominant use-cases detection (DU): Cluster extracted use-case phrases into semantically similar use-cases and merge clusters to detect dominant use-cases, and 3) Product mapping (PM): Associate each product in the product category with detected one or more dominant use-cases. Details of main components of the UBS system are provided below.

WSDM '24, March 4–8, 2024, Merida, Mexico

ACM ISBN 979-8-4007-0371-3/24/03.
https://doi.org/10.1145/3616855.3635748

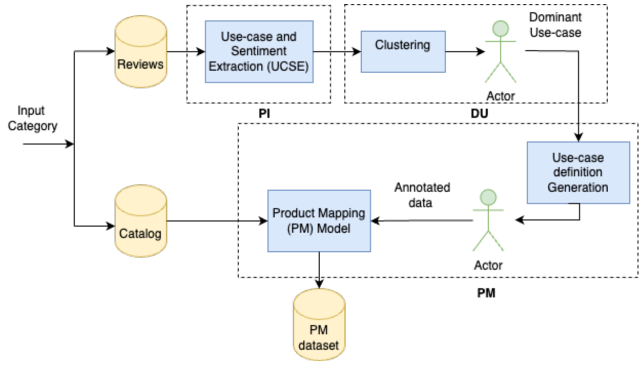

Figure 2: UBS ML Pipeline

2.1 Use-case phrase identification (PI)

The USB system takes category as input (e.g., laptop, refrigerator) and first runs PI on customer reviews for all products belonging to the input category. For this purpose, we train a joint use-case and corresponding sentiment extraction (USCE) model from customer reviews. We formulate UCSE as a generative task and fine-tune instruction tuned LLM [1, 2, 4] for it. Given a review text R, we prepend an Instruction Prompt (IP) to the review to obtain the modified input $IP(R)$. Here, IP contains a natural language description of the task. The target answer for the LLM therefore becomes: *<sentiment polarity> for <use-case-1> ; <sentiment polarity> <for use-case-2>* . . . , where "sentiment polarity" can be either positive, negative or neutral, and "use-case-i" is the i^{th} use-case phrase mention appearing in the review text R. Our primary objective in this task is to train LLMs to make them understand the definition of a use-case so that they can efficiently extract use-cases (and their corresponding sentiments) from reviews for unseen categories, thereby generalizing elegantly. We enhance the UCSE model with elaborate task-specific instructions, multi-task training, and finally incremental few-shot re-training to further improve the performance on new/unseen categories.

2.2 Dominant use-cases detection (DU)

Use-case phrases (e.g., online classes, student, school homework) extracted using the UCSE model are then clustered and used to identify the dominant use-cases (e.g., education) for the given product category (e.g., laptop). For use-case clustering, we use Claude v2, available through AWS Bedrock, to cluster semantically similar use-case phrases and automatically assign a name to the identified cluster. While the context length for Claude is large, there is a limit on the length of the generated output. Therefore, we developed an iterative prompting approach using Prompt Chaining. In each subsequent call to Claude, we ask it to identify clusters from a subset of use-case phrases. Further, we instruct Claude, not to repeat clusters that are already returned in previous responses. Also, for each coarse-grained cluster identified by Claude, we further design prompts to obtain a second level of fine-grained use-case clusters, thereby obtaining a 2-level hierarchical clusters of use-case phrases. These clusters are then validated with inputs from category and product managers to get dominant use-cases for a category.

2.3 Product mapping (PM)

Once dominant use-cases are defined for a category, definitions for use-cases are created before annotating the data for the PM. The annotated data is then used to train the PM model for mapping all the products for a given category to its dominant use-cases. Training category specific supervised models needs separate annotation data for each category, which is time consuming and expensive to obtain and is a major bottleneck in scaling model to thousands of product categories. Also, the learnings across categories are not transferable as models are trained independently and hence it's like training a model from scratch for every category. We formulate the problem of product mapping as product details to use-case definition matching task and harness this ability of LLMs to follow such instructions. We pose it as a Natural Language Inference (NLI) task where product details and use-case definition are sent as input to the LLM, and the hypothesis is to determine if the product matches the provided use-case definition. This formulation enables us to train a single model for PM mapping and also help generalize well on unseen categories, use-cases and even handle change in definition. We provide structured and unstructured product attributes as product details to the model and use standard operating procedure (SOP) created for annotators as the use-case definition.

3 COMPANY DESCRIPTION

Amazon.com, Inc. is a multinational technology company focusing on e-commerce, cloud computing, online advertising, digital streaming, and artificial intelligence. Founded in 1994, Amazon's business interests include online retailer and marketplace, cloud computing service through AWS, live-streaming service through Twitch. We cover search and information retrieval scenarios primarily focused in the field of e-commerce. Members of Amazon, appear frequently at information retrieval and knowledge discovery venues and have contributed significantly to these technologies.

4 SPEAKER BIOGRAPHY

Sachin Farfade is a Principal Applied Scientist with the International Machine Learning team at Amazon. Where he is leading science effort towards building ML solutions to provide navigational assistance to customers using Large Language Models. Before joining Amazon, Sachin worked at Yahoo! Labs, where he developed computer vision solutions using CNNs for Flickr and Yahoo! search. He received his Masters in Engineering from IISc, Bangalore in 2005. Sachin's research interest includes NLP and CV.

REFERENCES

[1] Hyung Won Chung, Le Hou, Shayne Longpre, Zoph, et al. 2022. Scaling instruction-finetuned language models. *arXiv preprint arXiv:2210.11416* (2022).
[2] Shayne Longpre, Le Hou, Tu Vu, et al. 2023. The Flan Collection: Designing Data and Methods for Effective Instruction Tuning. arXiv:2301.13688 [cs.AI]
[3] Long Ouyang, Wu, et al. 2022. Training language models to follow instructions with human feedback. http://arxiv.org/abs/2203.02155
[4] Colin Raffel, Shazeer, et al. 2020. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *J. Mach. Learn. Res.* 21, 1 (jan 2020).
[5] Hugo Touvron, Louis Martin, et al. 2023. Llama 2: Open Foundation and Fine-Tuned Chat Models. *ArXiv* abs/2307.09288 (2023).
[6] Wayne Xin Zhao, Kun Zhou, et al. 2023. A Survey of Large Language Models. *arXiv preprint arXiv:2303.18223* (2023). http://arxiv.org/abs/2303.18223