

ABSTRACT

With increasing internet accessibility, there is a rising interest in online shopping. The importance of e-commerce platforms has driven research in different directions like search, recommendation, attribute extraction, question-answering, etc. E-commerce portals process billions of queries daily, and query understanding is vital for accurate comprehension of search queries to return high-quality search results. Customers express their purchase intents in several ways, some of which may use a different vocabulary than that of the product catalog. Search engines typically suffer from poor performance on such queries because of the low overlap between query terms and specifications of the desired products. We introduce the Co-Attentive MaLSTM model that measures semantic similarity between a pair of queries to rewrite vocabulary gap queries to well-performing queries. Our technique considerably outperforms strong baselines, and online A/B experiment shows its effectiveness in real-world settings. To better understand how the query pairs are comparable, we chunk the sentences and focus on aligning the fragments. These alignments provide an explanatory layer over semantic textual similarity. We present a pointer network-based model with a sentinel gating function that uses ConceptNet and syntactic knowledge to align components of sentences.

Another pivotal component of e-commerce is the product attribute values or product specifications. Product attributes such as brand name, color, material, etc., are critical features that help in differentiating products. Attribute value extraction deals with extracting the values of attributes from the product profile. Unlike previous approaches that formulate the attribute value extraction as a Named Entity Recognition task or a Question Answering task, we propose to tackle the attribute value extraction

task using generative frameworks. We conduct experiments on two datasets where the generative approaches achieve new state-of-the-art results. We observe that a substantial fraction of user queries can be answered with product specifications. As manually annotating the dataset is time-consuming, we propose a semi-supervised method to create a large training dataset and leverage language models to answer the questions using the product specifications. Our models exhibit good performance even when trained on one vertical and tested across different verticals.

Apart from specifications, product reviews are a critical source of information to answer product-related questions. Since numerous questions are posted on the Question Answering platforms of e-commerce websites, there is a need to provide automated answers to questions. Retrieving relevant reviews is essential to generate natural answers to product questions. We use transformer-based review ranking models, which provide a ranked list of reviews as a potential answer to a new question. We also present a data augmentation technique to generate new questions from customer reviews. Experimental results show substantial improvements over the existing approaches using the data augmentation technique. Finally, we extensively analyze two state-of-the-art answer generation models in diverse scenarios to check the reliability of the generated responses. Our analysis shows that input reviews are not always utilized significantly for answer generation. Further, the models perform poorly in answering the numerical questions, and many generated answers contain unhelpful phrases. These shortcomings need to be addressed for a reliable generative product question-answering system.

Keywords: Vocabulary Gap; Product Search; Interpretable Semantic Textual Similarity; Attribute Value Extraction; Review Ranking; Product-Related Question Answering; E-commerce