Summary System for Amazon Appliances

Cristian Perez
University of Illinois Urbana Champaign
Urbana, Illinois, USA

Claire Zheng
University of Illinois Urbana Champaign
Urbana, Illinois, USA

Nolan Wang
University of Illinois Urbana-Champaign
Urbana, Illinois, USA

Yat Lam Chan
University of Illinois Urbana-Champaign
Urbana, Illinois, USA

Abstract

We plan to create a summary system that can help users gain insight on what to look for when shopping. The user would input a product, and the system would then summarize the features that users may want or look for within the product. To train the system, we plan to use the McAuley Lab Amazon Reviews dataset, focusing on the Appliance category. The system should cluster reviews, extract key words representing the cluster, and associate a rating to the cluster. The system returns key words for clusters that are clearly positive or negative. Evaluation would be done by checking if the output is reasonable, and time permitting, have users rate the output.

ACM Reference Format:
Cristian Perez, Claire Zheng, Nolan Wang, and Yat Lam Chan. 2025. Summary System for Amazon Appliances. In . ACM, New York, NY, USA, 1 page. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

When buying a product online, often choosing the cheapest or the most expensive ones may not yield a satisfactory experience. After using our summary system to search up a general product to learn about features they should consider, users can then access their favorite retail website to find the optimal product that would fit their specific needs. For example, if the user inputs "electric toothbrush" into our summary system, the output would be key features customers value highly for toothbrushes. We anticipate a sample output would mention toothbrush lifespan, cleaning power, different modes on the toothbrush, etc. We think that this would give users unfamiliar with a certain product category (but needing to use it) some baseline knowledge to help them better shop. Additionally, this could help sellers know what features they should be focusing on when designing a product and creating a listing for it.

2 Description of Intended Work

To solve this problem, we plan to use a clustering algorithm on the reviews, this way we can have a few key words from each review to represent a group. Then, we have the system learn which clusters are more associated with positive reviews and those that are associated with negative reviews. When the user enters a query, the system can then choose the clusters associated with highly positive and highly negative reviews and return the key words for those clusters to the user, so the user knows what to look for when reading reviews. The user would interact with the model through a web interface.

2.1 Datasets

We plan to use two datasets containing reviews and metadata from the category "Appliances" in Amazon, downloaded from https://amazon-reviews-2023.github.io/. User reviews include ratings, text, helpfulness votes, etc. about the product. The item metadata includes descriptions, price, raw image, etc. about the product. Reviews and metadata are connected through the parent ID of the product, which contains similar products that only differ in colors, styles, and sizes.

2.2 Evaluate and Test Software

After we are able to create a model that meets the requirements, we can check whether the model gives insightful and reasonable suggestions for a specific product group. For further analysis, we can also have students review the key words the system returns on a scale of helpfulness.

2.3 Brief Timeline of Work

We expect to spend 1-2 weeks on dataset processing. Then another 1-2 weeks on converting text to an appropriate vector representation. After that, 2-3 weeks on the clustering algorithm and generating the insights. Finally, another 1-2 weeks on the user interface. If we have additional time, we may try to get a larger sample of users and rating the key words the system returns.

2.4 Activities Per Team Member

Claire: Build a simple user interface for searching and retrieving summarized insights for products

Yat Lam Chan: Converting text to vectors to choose specific keywords/feature that matter for each product

Nolan: Data processing what features we want from the dataset, cleaning up those features

Cristian: Cluster products based on features or see if/which features correlated to review

Conference'17, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn