Topic Modeling and Sentiment Analysis on Amazon Reviews

Joan John and Daehan Kwak
School of Computer Science and Technology, Kean University, Union, NJ 07083

Keywords: e-commerce, online review, topic modeling, sentiment analysis, specific features

Due to the increased number of consumers using e-commerce platforms such as Amazon, there is a growing demand to offer consumers innovative services. Technological advancements have fostered changes in the e-commerce sector by prioritizing on the needs of the customers. For instance, e-commerce platforms allow customers to share opinions regarding a product through online reviews. These reviews have immense impact on a customer's decision to make a purchase because these opinions are based on another customer's experience with a particular product. There are innovative approaches used to ensure customers can access and read as many product reviews as possible to improve their purchase decision. Therefore, approaches like topic modeling and sentiment analysis can be used to discover the hidden semantic structures within the reviews.

Sentiment analysis involves drawing meaning from a text by using computer programs to understand and interpret paragraphs, and sentences by evaluating their structure and identify connections between words in a specific context (Ali et al., 2019). In the context of e-commerce, semantic analysis can be important in retrieving information and web data analysis due to the immense amount of user-generated content. This model can be used to extracting consumer opinions about products sold online. Topic modeling is a form of statistical application used to identify the abstract topics in a collection of texts or documents thus learning important patterns of words (Ali et al., 2017). When going through a product's reviews, some can have long feedbacks which can be difficult and tedious for other customers to pick out the precise opinion about the product. Therefore, semantic analysis and topic modeling come in to help consumers analyze a specific product by going through numerous sentiments within a short time thus enabling them to make a purchase decision faster and easily.

The objective of this research is to provide feature-specific ratings on products based on text reviews. For example, the feature-specific ratings for an Amazon's Echo Dot can be specific ratings for battery life, sound quality, design, etc. The specific features are extracted via topic modeling from the text reviews and the star ratings for the specific features are based on the sentiment of the text reviews. In this research, topic modeling and sentiment analysis are implemented in Python using the open-source APIs: Gensim and Textblob, respectively. Therefore, this study puts forward a feature-specific sentiment analysis for product reviews using an advanced natural language processing algorithm. The topic words that are extracted will be mapped, thus allowing feature-specific sentiment analysis on the product reviews. The study uses Amazon's Echo Dot dataset which is obtained from Kaggle because Amazon's platform has sufficient product reviews and related metadata that span a period. The dataset entails product metadata such as brand, price, and category information as well as reviews such as text, helpfulness votes, and ratings.

The study proposed is a state-of-the-art approach to foster feature-specific sentiment analysis that provides a convenient product review for customers instead of the conventional and unstructured review texts. Topic modeling was used on the dataset to unravel the hidden features of the product in text reviews hence grouping the data based on specific features and identifying the sentiment that characterized each of these text reviews. Topic modeling

together with sentiment analysis approach will evidence high accuracy for every aspect. As a result, this research will experimentally verify the effectiveness and importance of the proposed algorithm when it comes to extraction of feature-specific sentiments from dynamic and text datasets. Thereby, the discussed topic modeling approach can be used in feature extraction and feature-based sentiment analysis of unstructured product reviews expressed by customers.

Sentiment analysis involves extraction of subjective information from a language text and expresses the user's perspective towards a specific topic. Sentiments are feelings which entail emotions, opinions, and attitude thus making it a subjective impression. On e-commerce platforms, this approach is conducted through topic modeling. This approach entails the application of statistical methods, processing of natural language, and machine learning to identify, collect, and characterize the sentiment context of a particular text review about a product. Consequently, this approach will be useful for online shopping because customers can gain better product information and idea about its experience before making a purchase.

References

Ali, F., Kwak, D., Khan, P., El-Sappagh, S., Ali, A., Ullah, S., ... & Kwak, K. S. (2019). Transportation sentiment analysis using word embedding and ontology-based topic modeling. Knowledge-Based Systems, 174, 27-42.

Ali, F., Kwak, D., Khan, P., Islam, S. R., Kim, K. H., & Kwak, K. S. (2017). Fuzzy ontology-based sentiment analysis of transportation and city feature reviews for safe traveling. Transportation Research Part C: Emerging Technologies, 77, 33-48.