Generating Aspect-based Review Summaries Using ACOS Quadruples

Pooya Adami
padami@usc.edu

Xinyue Cui
xinyuecu@usc.edu

Victor Hui
victoryi@usc.edu

Saeedeh Mahmoodifar
saeedehm@usc.edu

Pinyi Wang
pinyiw@usc.edu

1 Project Domain & Goals

Recent advances in Aspect-Based Sentiment Analysis (ABSA) have shown great promise in extracting Aspect-Category-Opinion-Sentiment (ACOS) quadruples in opinionated content such as online reviews. However, most existing works focus on extracting opinions and sentiment on one review at a time. In this project, we explore different methods of generating summaries over multiple reviews of a product/restaurant/service using ACOS quadruples as well as multiple evaluation metrics to quantify the quality of the summaries.

Our team strongly believes that the project's outcome will have a positive impact on review sites such as Yelp or Amazon. Currently, the rating of a product is presented on a scale of 1 to 5, which we believe oversimplifies the nuanced and subjective nature of people's opinions. For example, a 1 star review might reflect negativity of a certain aspect of a restaurant, but others may not be concerned about the same aspect, which can unfairly decrease the overall rating when averaged together with other reviews. Our proposed review summarization system aims to provide context to ratings, empowering consumers with more information and freedom to make informed decisions, ultimately improve their overall satisfaction.

To achieve our goal, we utilize state-of-the-art ACOS quadruple extraction techniques with the fine-tuned T5 model introduced in (Peper and Wang, 2022). We group together reviews with the same category/aspect in their ACOS quadruples to deliver an aspect-oriented review summary. In order to retain the most important concepts (ACOS quadruples) during the summarization process, one approach is to augment the original input with ACOS quadruples. Alternatively, we could perform keyword based controlled text generation given ACOS quadruples.

The performance of our proposed model will be evaluated using standard evaluation metrics such as ROUGE, BLEU and BERTScore. Additionally, we extend metrics in (Bhaskar et al., 2022) to measure if our model is making accurate (faithfulness), specific (genericity), and factual (factuality) summarization.

2 Related Work

ABSA is a subfield of Natural Language Processing (NLP) that aims to identify and extract sentiments from a given text concerning specific aspects or attributes of a product or service (Hu and Liu, 2004). ABSA has gained significant attention from researchers and industry practitioners in recent years due to its potential applications in various domains, such as e-commerce, customer feedback analysis, and social media monitoring. One of the emerging tasks in ABSA is the extraction of ACOS quadruples, which not only identifies the aspect and sentiment but also associates them with a specific category.

Generative models have demonstrated impressive results in ABSA tasks, including ACOS extraction. Conditional Variational Autoencoder (CVAE) is a generative model that can generate reviews or summaries of products, services, or entities based on the input text (Li et al., 2018). Transformer-based models, such as BERT and GPT-2, have also been used for ABSA tasks, including ACOS extraction. Recent studies have shown that fine-tuning BERT on ACOS extraction tasks outperformed previous state-of-the-art models (Zhang et al., 2020). Variational Autoencoder (VAE) is another generative model used to generate informative reviews by conditioning on the input text and the aspect to be reviewed. VAE has shown promising results in generating high-quality reviews that reflect the sentiment and opinions expressed in the input text (Hu et al., 2019). Graph Convolutional Network (GCN) effectively captures the relationships between aspects, categories, opinions, and sentiments in the

text (Li et al., 2020).

Common ABSA subtasks include recognizing and categorizing aspect terms, identifying supporting opinion terms, and detecting sentiment polarity expressed in text. However, implicit language is present in over 30% of sentiment expressions (Cai et al., 2021), which makes it challenging for existing models to handle such cases accurately. Implicit aspects and opinions, such as "it took an hour to be seated," are particularly difficult for models to recognize since they lack explicit aspects and opinion terms that are typically used as indicators, such as "the service" and "slow." Hence, addressing the ACOS quadruple extraction task is crucial, particularly for formulations supporting implicit aspects and opinions.

Researchers have introduced generative models with techniques for improved structured generation for ACOS quadruple extraction. For example, Peper and colleagues have proposed a model to improve quadruple prediction by using supervised contrastive learning, which encourages the model to generate input representations that can be distinguished based on important input features, such as sentiment polarity and implicit opinions and aspects (Peper and Wang, 2022). However, these models focus on extracting ACOS quadruple on one review at a time. Therefore, we aim to examine various methods for generating summaries of multiple reviews based on ACOS quadruples resulting from the improved models. Additionally, we will explore evaluation metrics to quantify the quality of review summaries.

3 Datasets

We will utilize the following pre-existing datasets: Dataset from (Cai et al., 2021) paper, the dataset is structured such that every line contains a review text and its corresponding quadruples. Each quadruple is formatted as 'Aspect Category Sentiment Opinion'.[1]The next dataset is a smaller portion of Yelp's collection of businesses, reviews, and user data.[2] As part of our data cleaning process, we will be specifically retaining only those reviews written in the English language. We will test several cleaning procedures, including but not limited to converting all reviews into lowercase, removing any HTML and URLs present, executing contractions, and tokenizing the review bodies. We aim to measure the effectiveness of these procedures. In order to further evaluate the efficacy of our model, we intend to conduct testing on the following datasets in our future work: Dataset from Amazon reviews. The dataset contains several million customer reviews of Amazon products as input text, along with star ratings assigned to each review as output labels.[3] Also the teaching evaluation dataset for professors was sourced from RateMyProfessor.com.[4]

4 Technical Challenge

As training the ACOS quadruple extraction technique detailed in (Peper and Wang, 2022) requires the T5-Large model with 770M parameters. The primary technical challenge we face is our lack of specific domain knowledge, computation resources, and, most importantly, time. To overcome this challenge, we intend to use their pre-trained model for restaurant reviews, though this is subject to further testing to confirm the model's feasibility with our selected dataset. Fortunately, several of our team members have access to USC's High Performance Clusters, which can be leveraged to mitigate issues regarding computation resources, providing us with a contingency plan in case we encounter issues with the pre-trained model.

Furthermore, extensive investigation is necessary to fully comprehend the field of Aspect based Sentiment Analysis as this domain is entirely beyond the scope of the CSCI 544 curriculum. However, due to the unique combination of our team's research and Machine Learning expertise, we are confident in our ability to achieve our objective.

5 Division of Labor

To foster a collaborative work environment, we have assigned overseeing responsibilities for each person in the group over specific portions of the project: Dataset + Pre-processing (Saeedeh), Paper Recreation (Victor), Opinion summarization research (Xinyue), Implement ACOS quadruples (Pinyi), Evaluation (Pooya). We are confident that our structured approach will lead to a positive experience for the team.

[1] https://github.com/NUSTM/ACOS/tree/main/data

[2] https://www.kaggle.com/datasets/yelp-dataset/yelp-dataset

[3] https://www.kaggle.com/datasets/bittlingmayer/amazonreviews?resource=download

[4] https://data.mendeley.com/datasets/fvtfjyvw7d

References

Adithya Bhaskar, Alexander R. Fabbri, and Greg Durrett. 2022. Zero-shot opinion summarization with gpt-3. *ArXiv*, abs/2211.15914.

Hongjie Cai, Rui Xia, and Jianfei Yu. 2021. Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 340–350.

M. Hu and B. Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*.

Z. Hu, Y. Cheng, N. Peng, and X. Huang. 2019. Improve aspect sentiment classification with aspect-oriented variational autoencoders. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*.

P. Li, L. Huang, J. Liu, and L. Wu. 2020. A graph-based framework for aspect category opinion target extraction. In *Information Sciences*.

Y. Li, H. Chen, J. Yang, T. Bai, and X Zhang. 2018. Conditional variational autoencoder for sentiment analysis of social media text. In *IEEE Access*.

Joseph Peper and Lu Wang. 2022. Generative aspect-based sentiment analysis with contrastive learning and expressive structure. In *Conference on Empirical Methods in Natural Language Processing*.

Y. Zhang, J. Yang, and D. Zhao. 2020. A survey on aspect-based sentiment analysis. In *Journal of Big Data*.