Fine-Tuning BERT for Aspect Extraction in Multi-domain ABSA

Arwa Akram, Aliea Sabir
Faculty of Computer Science and Information Technology, Computer Information System Dept., University of Basrah, Basrah, Iraq

E-mail: itpg.arwa.akram@uobasrah.edu.iq, aliea.sabir@uobasrah.edu.iq

Aspect extraction plays a crucial role in understanding the fine-grained nuances of text data, allowing businesses and researchers to gain deeper insights into customer opinions, sentiment distributions, and preferences. This study presents a BERT-based framework for aspect extraction in ABSA and evaluates its performance. Our research focuses on the comprehensive analysis of aspect extraction as we test our method using the SEMEVAL dataset of various consumer evaluations across diverse domains, including laptops, restaurants, and Twitter. By fine-tuning BERT on a large dataset, we aim to overcome the limitations of traditional approaches and improve the accuracy and efficiency of aspect extraction in ABSA. The experimental findings provide evidence of the efficacy of our methodology with a noteworthy aspect extraction accuracy of 0.98, highlighting its capacity to properly and consistently extract features. The article also explores the applicability of our approach to new domains and its possible applications in real-world scenarios.

Povzetek: Ta študija predstavlja okvir za ugotavljanje aspekta v ABSA, ki temelji na sistemu BERT.

1 Introduction

Aspect extraction, also known as feature extraction is a fundamental task in natural language processing (NLP) that involves identifying and extracting specific aspects or features from text data. These aspects refer to the different elements, attributes, or components that are being discussed or evaluated within a given context. The background of aspect extraction can be traced back to the broader field of sentiment analysis, which aims to understand and interpret subjective information, such as opinions, sentiments, and emotions, expressed in text. While sentiment analysis provides an overall sentiment polarity (positive, negative, neutral) associated with a piece of text, aspect extraction goes further by identifying the specific aspects or features that contribute to the expressed sentiment. Aspect-Based Sentiment Analysis (ABSA) is an essential subdomain of natural language processing (NLP) that concentrates on the extraction and analysis of particular aspects included in textual material [1]. This approach enables a deeper understanding of the attitudes expressed towards these specific features. The primary obstacle in ABSA (Aspect-Based Sentiment Analysis) is to the accurate detection and extraction of aspects, which serves as the fundamental basis for sentiment analysis at the aspect level. This work's central focus is optimizing aspect extraction in the field of Aspect-Based Sentiment Analysis (ABSA), with a specific emphasis on situations that include multiple domains. Although ABSA has made significant advancements in recent years, the extraction of aspects continues to be a complex and context-dependent task. Conventional techniques for aspect extraction were labor-intensive and frequently failed to capture the underlying complexity and diversity of language since they mostly depended on manually created features and domain-specific rules. Considerable progress has been made in natural language processing tasks such as sentiment analysis and aspect extraction with the development of deep learning models such as BERT [2]. The main goal of this work is to optimize the BERT algorithm's capacity for aspect extraction in consumer reviews. BERT is a pre-trained transformer-based language model that has shown impressive performance on a range of natural language processing tasks by making use of its attention mechanisms and contextual embeddings. Our goal is to overcome the drawbacks of existing methods and improve aspect extraction accuracy and efficiency in Aspect-Based Sentiment Analysis (ABSA) by fine-tuning BERT on a large dataset of customers reviews. The challenge lies in the identification and isolation of elements within textual data that are not only related to items of interest, such as computers, restaurants, and Twitter debates but