Project proposal: Unmasking Sarcasm - Enhancing Sentiment Analysis in E-Commerce Reviews and Questions

Hugo Bouy
Illinois Tech
hbouy@hawk.iit.edu

Rémi Kalbe
Illinois Tech
rkalbe@hawk.iit.edu

Mathias Roumane
Illinois Tech
mroumane@hawk.iit.edu

1. Problem description

Online reviews and questions on e-commerce platforms like Amazon play a critical role in guiding consumer purchasing decisions. However, sarcasm and humor in reviews and questions can mislead sentiment analysis tools, affecting both sellers and buyers. Our project aims to develop a machine learning model to detect sarcasm and humor in Amazon reviews and questions, ensuring a more accurate representation of consumer sentiment.

2. Brief Survey of Previous Work

In the work by Poria et al. [**?**], the authors present a sarcasm detection method using deep convolutional neural networks (CNNs). They argue that traditional approaches, which treat sarcasm detection as text categorization, often miss the deeper understanding of language nuances required for sarcasm. Their method integrates sentiment, emotion, and personality features extracted from pre-trained CNNs. By leveraging Twitter data, they contrast sarcastic sentences with the ground-truth polarity of events, and their experiments with word embeddings from word2vec and a combined CNN-SVM scheme showed superior performance on benchmark datasets.

Jain et al. [**?**] explored the complexities of identifying sarcasm in Amazon reviews. Recognizing sarcasm is crucial for accurate sentiment analysis, especially since sarcastic comments can be misinterpreted by traditional opinion mining methods. They utilized the "Sarcasm Corpus" containing labeled ironic and regular Amazon reviews, extracting features like sentiment scores, punctuation patterns, and contextual elements that consider the contrast between review sentiment and product rating. Their experiments showed the Support Vector Machine (SVM) classifier as the most accurate, emphasizing the role of context in sarcasm detection.

Yaghoobian et al. [**?**] discussed the challenges of sarcasm detection in sentiment analysis. They identified shifts in sarcasm detection methodologies and categorized detection methods into content-based, which focus on lexical indicators, and context-based, which emphasize background knowledge. Their study highlighted the CASCADE model, which uses user embeddings to capture user-specific features, as an example of leveraging context for sarcasm detection.

Ziser et al. [**?**] demonstrated product bias in Product Question Answering (PQA) systems, where certain products attract more humorous questions. They proposed a deep-learning framework to detect humor in PQA, focusing on incongruity and subjectivity.

Annamoradnejad and Zoghi [**?**] proposed a humor detection model named ColBERT, which uses BERT embeddings for sentence representation, achieving state-of-the-art results on various datasets.

Gupta et al. [**?**] explored the use of Large Language Models (LLMs) in humor detection, emphasizing their capability to capture the intricacies associated with humor and offense detection.

References