Analysis of languages used in online reviews

Isabelle Voskamp, S5319242

Abstract

This research focuses on the analysis of language used in online reviews with the goal of determining which words contribute to positive or negative sentiments. The research will conduct a comprehensive analysis of online reviews on the Internet with a focus on words that influence sentiments. This research will not only contribute to a better understanding of language use in online reviews, but it will also be relevant to discussions of sentiment analysis and consumer behavior.

1 Introduction

In recent years, the rise of online reviews has played a role in shaping consumer choices and influencing the reputation of products and services. Online reviews have become widespread in the digital age and impact customers' choices and ways of thinking.

This research paper will focus on specific component of online reviews, namely the influence of specific words in forming positive or negative sentiments. The purpose of my research is to examine the relationship between the use of language in online reviews and the resulting sentiment. The question under discussion is: *how do specific words in online movie reviews contribute to positive or negative sentiments?* The platform where I will examine movie reviews is The Internet Movie Database (IMDb).

Online reviews are based on credibility (credible or not credible) and sentiment (positive and negative). This provides an analysis that examines the relationship between language, credibility, sentiment and consumer decision-making.

I expect that differences in the language used in online customer reviews evoke different sentiments and will influence consumers in making decisions. This expectation is based on the idea that the language used in online reviews directly affects customers' feelings, influencing their choices. In my opinion, it is crucial to investigate language variations in online customer reviews to better understand how customers build trust, experience satisfaction and make choices.

The motivation behind this research is based on the recognition of the important role language plays in online consumer environments. For companies striving to improve consumer trust, satisfaction, and decision-making, understanding the factors that contribute to sentiment in online reviews is critical. By examining the differences in language, we seek to understand how these elements interact and how companies can develop strategies to improve their online reputation. This research provides useful insights that can be used to better understand and influence how customers interact with online reviews.

	Credible reviews	Non-credible reviews
Positive sentiment		
Negative sentiment		

Table 1: Table 1: This provides an analysis that examines the relationship between language, credibility, sentiment and consumer decision-making.

2 Method

To begin the research, I carefully determined the search terms for the online reviews in order to gather targeted information that is actually valuable. The focus is mainly on the most recent online reviews on the platform IMDb, using a similar length of time to optimize posts.

IMDb is an online database of information about movies, television shows, actors, directors

and other related content. People around the world use IMDb, a comprehensive source of information about movies and television, to find details about their favorite movies and TV shows. It is possible for users to leave ratings and reviews for movies and TV shows on IMDb, which helps them choose what to watch. This approach contributes to the consistency of the results and ensures that the online reviews are relevant because they are, after all, current.

The application of sentiment analysis is an important aspect of this research. Sentiment analysis is the technique of identifying important and individual data from a collection of source material that contains unrecognized information about people's perspectives. (Giatsoglou et al. 2017). To automate the organization and categorization of sentiments from reviews, sentiment analysis uses text analysis, natural language processing (NLP) and computational techniques. (Hussein 2018).

The goal of sentiment analysis is to investigate and determine users' attitudes toward different domains, i.e., "negative or positive." (Shaukat et al. 2020).

I have committed to using specific terms on the platform, such as words that express a positive feeling in online reviews, such as "excellent," "great," "fantastic," "perfect," "pleasant," "satisfied," and "recommend. I also use words that express a negative feeling in online reviews, such as "bad," "disappointing," "terrible," "difficult," "unacceptable," "unreliable," and "unpleasant. By consciously choosing those words, I try to find words with which to compare the reviews. By doing so, I am trying to determine whether the reviews on IMDb that I have found are positive or negative with the search terms used. The results of this analysis will not only indicate which emotions are prominent in the reviews, but they will also serve as an example for the results of my research.

By proceeding in this way, I hope to not only gain a better understanding of how certain words in online reviews affect feelings, but also to learn more about the factors that influence consumers in making decisions.

3 Related work

Research 1

Shaukat et al.'s research focuses on sentiment analysis of IMDb movie reviews. To categorize sentiments, they use lexicon-based techniques and neural networks. This differs from my research, which focuses on selecting carefully chosen keywords for the most recent online reviews on IMDb. (Shaukat et al. 2020).

My research focuses on using IMDb as a useful source of information about movies and television shows. I ensure that search terms are optimized to gather current, targeted information. In my research, I emphasize the importance of IMDb as a globally known database where users leave reviews to help others choose movies.

Both studies share a common element of sentiment analysis. Text analysis, natural language processing and computational techniques are used to understand people's opinions. But by using specific positive and negative words, I focus on organizing and categorizing feelings from reviews. This provides an understanding of the emotions that stand out in reviews and illustrates the findings of my research.

My use of specific words in online reviews, such as "excellent" and "poor," to determine users' opinions highlights the methodological diversity between the two studies. This constitutes an alternative method compared to the use of WordNet dictionaries in other studies.

By carefully choosing particular words and examining how they influence emotions in reviews, one gains a better understanding of emotions and the factors that influence consumers in making decisions. This improves my knowledge of the links between certain words, emotions and consumer behavior in the context of IMDb reviews.

Research 2

Prashuna Sai Surya Vishwitha Domadula et al.'s research analyzes sentiments in movie reviews on IMDb by using both a lexicon-based approach and the BERT neural network model. The goal is to convert textual reviews into comprehensible data for machine learning classifiers. It identifies key terms such as "sentiment analysis," "opinion mining" and "neural networks" and evaluates the lexicon-based method and the BERT model based on performance evaluation metrics. (Domadula and Sayyaparaju 2023)

My research approaches the online reviews on IMDb with carefully chosen search terms, unlike this study. I focus mainly on the most recent reviews and emphasize the importance of IMDb as a

globally recognized source for information about movies and television. The focus is on understanding the emotions reflected in these reviews by using specific positive and negative terms.

By focusing on certain words and how they affect feelings, it can help us understand more about how customers make decisions. My analysis results serve as an example for the larger framework of my research on sentiments on IMDb.

Research 3

Mohammed Mohaiminul et al.'s research focuses on how sentence-level phrases are classified in IMDb movie reviews. They chose a two-part classification system, where words are classified as positive or negative. The IMDb dataset with 25,000 reviews (12,500 positive and 12,500 negative) gives the model a rich source of information. (Islam and Sultana 2018)

The Bag of Words (BoW) model with feature selection based on unigrams is an important part of their research. Data processing and natural language processing use the Python machine learning library. Developing a sentiment analysis process that compares feedback processing is part of the research process. Figure 1 shows the workflow used for sentiment analysis of text and comparing review processing.

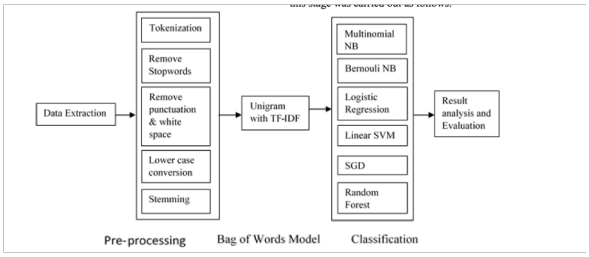

Figure 1: Workflow model for review processing.

The research extracts data and selects only relevant fields to maximize memory usage. By removing stop words and splitting sentences into unigrams, features can be extracted. Words are converted into integer values by the feature vector. The classification procedure involves assessing a number of methods with a 5-fold cross-validation method, using ratings and review texts as features.

By evaluating the effectiveness of the BoW approach and comparing it to alternative methods, the research positions itself within a broader framework of sentiment analysis. It improves the understanding of sentiment analysis at the sentence level and highlights the methodological subtleties associated with the use of different classification models.

In contrast, my research focuses on keywords selected from the most recent online reviews on IMDb, analyzing both positive and negative aspects of the search terms. My research focuses on how certain words influence feelings, but the research by Mohaiminul et al. looks at phrase classification techniques in a broader context of sentiment analysis, according to IMDb. Both studies provide unique insights into the emotional dynamics of IMDb reviews.

4 Predicted Results & discussion

The results of the study of specific positive and negative terms in recent IMDb reviews shed light on the emotional dynamics of user content on the platform. Identifying crucial words that evoke emotions plays an essential role in understanding how users express their opinions on IMDb.

The findings provide insight into how certain words shape overall sentiment in reviews, with the study focusing on words such as "excellent," "disappointing," "fantastic" and "terrible. Compared to other sentiment analysis approaches, my research emphasizes methodological diversity by focusing on carefully chosen search terms in recent reviews. Although lexicon-based techniques and advanced neural network models are valuable, this targeted word selection offers an alternative approach.

5 Conclusion

By using sentiment analysis, with a focus on carefully chosen positive and negative words, this study sought to discover emotional aspects in recent IMDb reviews. The goal was to discover how certain words influence users' feelings by highlighting the aspects that influence their opinions.

The analysis showed that terms such as "excellent," "fantastic" and "perfect" contributed strongly to positive sentiment in IMDb reviews, with people who used these terms being satisfied and appreciative. On the other hand, terms such as "disappointing," "terrible" and "unacceptable" were strongly associated with negative emotions, causing users to express dissatisfaction and dissatisfaction with the reviewed content.

The method of targeted search terms and more developed sentiment analysis methods highlighted

the diversity of the research. The research acknowledges limitations, such as the lack of contextual depth in the reviews studied and the limited generalizability of the results.

In conclusion, this study not only provides a great deal of information about the emotional aspects of IMDb reviews, but it also highlights how important improvement techniques are. The goal is to gain a better understanding of how users experience digital content, with room for further refinement and expansion of research areas.

References

Domadula, Prashuna Sai Surya Vishwitha and Sai Sumanwita Sayyaparaju (2023). *Sentiment analysis of IMDB movie reviews: a comparative study of Lexicon based approach and BERT Neural Network model*. URL: https://www.diva-%20portal.org/smash/get/diva2:1779708/FULLTEXT02.pdf.

Giatsoglou, Maria et al. (2017). "Sentiment analysis leveraging emotions and word embeddings". In: *Expert Systems with Applications* 69, pp. 214–224.

Hussein, Doaa Mohey El-Din Mohamed (2018). "A survey on sentiment analysis challenges". In: *Journal of King Saud University-Engineering Sciences* 30.4, pp. 330–338.

Islam, Mohammad Mohaiminul and Naznin Sultana (2018). "Comparative study on machine learning algorithms for sentiment classification". In: *International Journal of Computer Applications* 182.21, pp. 1–7.

Shaukat, Zeeshan et al. (2020). "Sentiment analysis on IMDB using lexicon and neural networks". In: *SN Applied Sciences* 2, pp. 1–10.