AAI-501-02-SU23 - Final Project Proposal: Analyzing McDonald's Customer Reviews Through NLP and Machine Learning Techniques

Paul Parks, Lishi Wang, Ivan Steklov

University of San Diego

AAI-501-02-SU23 - Introduction to AI

Ying Lin

July 17, 2023

Abstract

This proposal outlines our final project to extract actionable insights from a large dataset of over 33,000 McDonald's Google reviews using advanced Natural Language Processing (NLP) and machine learning techniques. The objectives of the proposed project include conducting sentiment analysis on customer reviews, exploring geographic trends in reviews and ratings, and predicting star ratings based on review text. We will use and assess several machine learning models using suitable metrics, such as linear regression. The proposal emphasizes the importance of data cleaning and model selection. If successful, the study can be used for businesses to leverage NLP and machine learning to enhance customer satisfaction, user reviews, business growth, resource management and operational efficiency.

Related Course Topics

This project will incorporate several related course topics, including:

- NLP (Natural Language Processing): Cleaning, preprocessing, vectorization, and embeddings.
- Deep Learning: Applying models for sentiment analysis.
- Supervised Learning: Implementing regression techniques to predict review ratings, possibly via Bayesian Networks.

Expected System Behaviors

The system is anticipated to demonstrate the following behaviors:

- Efficiently classify review sentiments as positive, negative, or neutral.
- Identify potential geographic trends in review sentiments or star ratings.
- Predict the star rating based on the review text with reasonable accuracy.
- Extract common sentiments, phrases and keywords.

Focus Areas

We will concentrate on the following key areas: data cleaning and text preprocessing, model selection and tuning, and interpreting results for sentiment and geographical patterns. The focus of the model training and analysis will be on sentiment analysis of the reviews. From there we will fit these sentiments into a gradient from positive to negative. We will also conduct analysis of certain context, repeated phrases and common sentiments grouped by region. This can inform us if there are any common issues within a region of Mcdonald's restaurants.

Resources we intend to use

Russell, S. & Norvig, P. (2021). Artificial Intelligence, A Modern Approach (4th). Pearson.

- Chapter 21: Sections 21.1.1, 21.1.2, 21.2.3, 21.4, 21.5, 21.6, and 21.8
- Chapter 23: Sections 23.1, 23.1.1, 23.1.5, 23.2, 23.3, 23.4.2, and 23.6
- Chapter 24: Sections 24.2

Hariharan, K. et al. (2023). Analyzation of sentiments of product reviews using Natural Language Processing. *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 1851–1854. doi:10.1109/icaccs57279.2023.10112777.

Hoque, M. M., Mahbub, M. S., & Al-Mamun, A. S. M. (2008). Isolating significant phrases in common natural language queries to databases. *2008 11th International Conference on Computer and Information Technology*, 554–559. doi.org/10.1109/iccitechn.2008.4803022