Opinion Mining with Sentiment Analysis

Using Google Chrome Extension

Group Members:

S. Barath, 1940225

Hyeok Kim, 1940214

Pranav Gopalkrishna, 1940223

Guided by:

Assistant Professor Roseline Mary

Sentiment Analysis Google Chrome Extension

Introduction

Sentiment analysis is a natural language processing (NPL) technique used to determine the underlying subjective sentiment (usually divided into the categories "positive", "negative" and "neutral") behind certain data. Sentiment analysis is often performed on textual data, and can be used to:

- Monitor brand and product sentiments in customer feedback
- Understand customer needs and desires

Challenges

Natural language processing (which will be critical in sentiment analysis) is tricky due to the dependence of word meanings on context. Many words take different meanings and convey different sentiments in different contexts. Creating a Google Chrome extension is relatively straightforward, but to make it capable of collecting the textual data (present in reviews and feedback) and store this data in a usable backend will be challenging.

Also, opinion mining deals with unstructured data in the form of text such as comments and blogs posted on the internet. Text dta has difficulty in analyzing because the meaning varies depending on adjectives and adverbs, or because there are many relative expressions such as "big" and "small", and undefined expressions. Therefore it is necessary to establish sentimental polarity and sentimental lexicon of vocabulary that objectively quantifies and statistics this.

Objective

The objective of this project is to create a user-friendly Google Chrome extension that can perform sentiment analysis on the reviews and feedback present in any webpage. In particular, we will aim for fine-grained and aspect-based sentiment analysis, but may also include emotion sentiment analysis.

Advantages of the proposed project

It is time-saving and efficient, as it can gather and process a lot of data and give us reliable information about the different aspects of textual data, including subjective aspects. We can use it to understand the nature of the texts, identify feedback with particular characteristics, or for particular aspects or purposes. This software can also be used to safeguard or prevent mishaps such as the spread of fraudulent messages or reviews, or highly negative responses about certain key areas.

As a student, we can use, apply and as well as learn a lot of concepts in this as this uses regression, artificial intelligence, classification, machine learning and data analytics..

Chrome Extension

Chrome extensions are mini versions of software programs that allow us to add and perform various functionalities in the browser, with minimal interface. It is built using web technologies such as HTML, CSS, JavaScript. The main purpose of it is that it allows us to add or customize our own features for any web page by accessing various JavaScript APIs. Mainly, It needs a manifest file in Json which is to be created that contains basic details like name, description, version number etc, which will be used to manage the app.

Implementation details, frontend & backend, and features

Backend (including database operations and sentiment analysis query functions) will be done using Python, and the web framework for this project will be developed using Django. The frontend development is undecided for now. Expected features for our software are:

- Public accessibility
- Fine-grained analysis (with predefined options)
- Aspect based analysis (with predefined options)
- Emotion analysis (fixed classes)
- Applicability to a variety of websites containing reviews, feedback or other textual data

Module details

Data gathering module

Organized data collection of the textual data is the first step in the analysis, and requires:
- Web scraping to automatically draw and organize data from the given website
- Backend that can store and retrieve data for the sentiment analysis Chrome extension

Analysis modules

There are expected to be three analysis modules, one for fine-grained analysis, one for emotion analysis, and one for aspect-based analysis (each having their own trained models and methods).

Tools & technologies used

- Chrome developer tools
- Python
- Natural language processing (NLP) (part of machine learning) (using TensorFlow)
- Data storage, processing and analysis (using Pandas)
- Web framework (using Django)
- Prediction (of sentiment, based on trained models)

Users & clients for the application

- Businesses looking for reviews on their products and responses to their campaigns
- Application developers looking for feedback on their application
- Surveyors
- Market researchers

Assumptions

We are assuming the following for our project:

1.

We will be able to read and organise textual data from any website, and use this data within our analytical code at runtime.

2.

We will be able to train models for recognizing aspects (particular topics of interest) effectively for a wide range of applications, i.e. we will have sufficient and accurate training data.