

Capstone

Problem Statement

Sentiment Analysis and NLP

Matthew E. Miller, Manisha Kumari, Lucki Ratsavong
9-12-2019

Project Description:

The goal of this project is to create business value through the use of sentiment analysis and natural language processing (NLP). Sentiment analysis is the process of surmising context, tone, and meaning from written text. It's a sub-discipline of natural language processing (NLP) and it's a powerful tool in a world with increasingly digital text.

Sentiment analysis processing can be performed on any set of digital text. For this reason, we can reasonably extend its reach by using auxiliary technologies to bridge the gap between non-digital text and the digital world by translating non-digital data into a digital form. For example, optical character recognition (OCR) can be used to process images of printed, hand-written text, into digital text. This data can then be processed with sentiment analysis and NLP tools, allowing us to quickly quantify historical documents/data.

Another possibility is the transcription of audio from real-time telephone conversations. First, audio transcription technologies must be used to translate the audio into digital text. Then, this digital text can be processed with sentiment analysis and NLP tools. For example, this might be used to evaluate the performance of a call-center employee.

As you can image, there are many forms of data which can be analyzed with NLP and sentiment analysis. This project is about processing these data sets and providing meaningful feedback to our customers and users.

Problem Description:

Currently, our customers and users are manually interpreting written text from various sources like course evaluations, customer surveys, twitter mentions, etc. The problem is that manual interpretation is slow, unquantified, and prone to discrepancies in understanding between readers. Our goal is provide automatic, programmatic, consistent, and quantifiable interpretations of written text. The outcome is that our customers/users gain better understanding, stronger analytical tools, and an improved ability to provide high-quality service in their professional roles.

We want to make better, more responsive, data-driven organizations and professionals.

Project Objectives:

- Build a functional sentiment analysis and NLP web application.
- Provide meaningful and accurate sentiment analysis and NLP results for customers and users.
- Improve upon existing sentiment analysis and NLP technologies.

Customer Identification:

Customer 1	Customer 2
Name: Lindsey M. Myers (Uni. Marketing) Objectives: Gain meaningful insight from customer feedback. Stakeholders: Employer (KSU), Clients	Name: Dr. Khan (CS Department Chair) Objectives: Improve quality of course offerings and the department. Stakeholders: Employer (KSU), Faculty
Customer 3	Customer 4
Name: Dr. Tang (Journalism and Comm.) Objectives: Unspecified. Stakeholders: Unspecified.	Name: Kent State University Objectives: Improve students' experience, course offerings, and quality of course offerings. Stakeholders: The State, Board of Directors, Customers (students), Employees

User Identification:

Our users are researchers, managers, and other people that want to quantify, previously, unquantified digital text. This pool of users is large, but we have chosen to focus our project on a few target users (our customers). These users typically have a background that includes higher education, management work, critical thinking, and sophisticated use of technology. These users likely work in areas that receive customer feedback which would normally be manually interpreted and possibly quantified using many manual man-hours.

Customers’ Constraints:

In some cases, the customer’s data is sensitive and cannot be released to aid in the software product’s development. The result is that mock data must be manufactured to accurately reflect the characteristics of an authentic dataset. This is a major constraint on development.

The other customer-constraints are schedule and resource based. The project must be completed within three months and programming work must be done within two. Furthermore, our customers are not paying for machine resources, so we must abide by the free-tier-restrictions and limit CPU time to 100s per day, store only 512MB on the remote server, and adapt to a low bandwidth environment.

Risks and Assumptions:

Assumptions	Risks
They care to improve services. They want to quantify written comments, reviews, etc. Quantifiable data is more valuable than unquantifiable, human-interpreted, information.	Low accuracy of the sentiment analysis. Poor AI modeling. Product may not be valuable, useful, usable for customers/users.

Conclusion:

We are building a web-based sentiment analysis and NLP tool to enable our customers and users to quantify previously unquantified text.