Checkpoint 5: Natural Language Processing
Findings Report

The Creative Wolves
Masum Patel, Sizheng Zhang, Jiawei Zhang

Introduction

Apart from the attributes that we have discovered in the previous four checkpoints, we hoped to identify more subtle differences between the Repeaters and the Offenders group in the Chicago Police Department. Currently, we know that Repeaters tend to have higher salaries, more awards, and have more frequent deployments in the Southern side of Chicago. In Checkpoint five, we utilized the power of natural language processing techniques to investigate patterns that lay in the allegation summary Repeaters and Offenders. More specifically, we used Vader sentiment analysis to analyze the report summaries that belong to each group respectively.

Procedure

The goal of this checkpoint is to understand whether there is a difference between the sentiment score of the allegation summaries that belong to Repeaters and those belonging to Offenders. We used Vader sentiment score library and the Natural Language Toolkit (NLTK) python library to complete our analysis. Our expectation was that Repeaters should have a noticeably more negative sentiment score while the Offenders would have slight better sentiment in their allegation reports. We moved away from the original checkpoint proposal due to the lack of textual evidence related to the findings we had in the previous checkpoints (information about salary, awards, or location); as a result, we decided to expand from our project theme and use NLP techniques to further identify more common attributes of Repeaters. We believe that more critical or negative comments can certainly be yet another distinguishing feature of Repeaters.

We first export two csv files from Datagrip, each representing the allegation summary samples belonging to each group. Since the Repeaters group consisted of 4000 officers with the most allegation count, we matched 824 allegation reports linked to these officers that have valid summary data. However, by taking a random sample of 4000 offenders, we only matched about 350 allegation reports linked to these officer_ids. As a result, we took a random sample of 824 allegation reports and compared them to the 824 reports of the Repeater group in order to ensure a fair comparison and more intuitive result.

Once we had the data ready, we loaded the csv into dataframes using Pandas and tokenized the sentences into tokens using the NLTK library. We then filtered the stopwords and punctuations from the sentences in order to reduce noise. Then each sentence goes through the Vader sentiment analyzer and receives a sentiment result that includes how positive/negative the sentence is and an overall compound score. Based on the Vader library, a compound score larger than 0.05 signifies a positive sentence and scores lower than -0.05 signifies a negative sentence; anything in between will be classified as neutral. We then count

the number of positive sentences and negative sentences from the samples of each group and compare their results.

Findings

We performed two sets of sentiment analysis and the following figures (1 and 2) show the result for 824 Offenders Allegation Reports and 824 Repeaters Allegation Reports. Note that the Positive and Negative scores are calculated by averaging each sentence's positive or negative score. The score measures how positive or negative each sentence is and is used to calculate the final compound score of each sentence, which categorizes the sentence into either a positive one or negative one.

Based on the result, the 824 Offender allegation summaries contain 751 negative sentences or 91.14% of all sample sentences while the Repeater summaries contain 785 negative sentences or 95.27% of all sentences. While the difference between negative sentence count is low, the difference in scores and percentage falls between our expectations. Moreover, the Repeater summaries have a higher negative score--meaning that the summaries themselves contain more negative or critical English words. Also note that the Repeater summaries have only about half of the positive summaries when compared to Offender summaries. While it is true that allegation summaries are supposed to be negative comments on the officer, the difference in positive and negative sentences among the two groups are evident and worth-noticing.

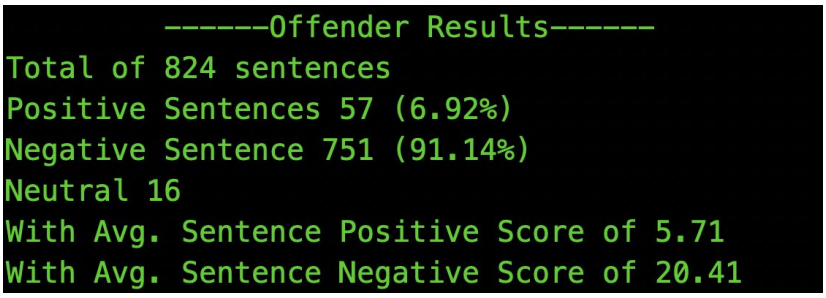

Figure 1: Sentiment Analysis Result of Offenders Allegation Report

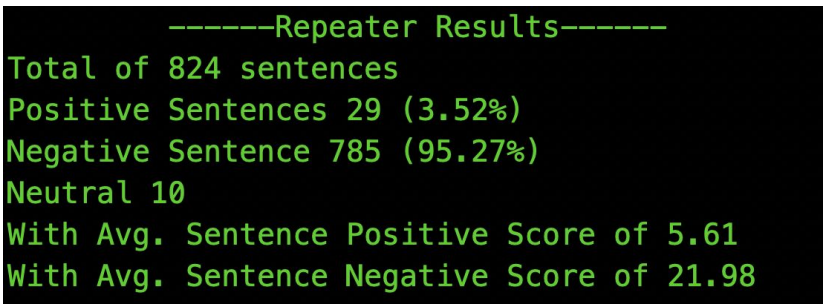

Figure 2: Sentiment Analysis Result of Offenders Allegation Report

Conclusion

From our analysis, we believe that Repeaters, the top 4000 officers who received the most allegations in the force, tend to have more negative allegation summaries in their report than general officers who have received at least one allegation. Based on the results returned from the Vader sentiment analysis, Repeaters also tend to have more negative comments in their summaries and much less positive contents when compared to the overall population of officers who have received at least one allegation. As a result, NLP has helped us to identify another common attributes of the Repeaters. While the difference between Repeater and Offender is small, the result makes logical sense and coincides with our expectations.

The result also opened a lot of opportunities for further investigation, since the sentiment analysis offered a direction. In the future, we believe it is worth diving deeper into the meanings and semantics of the allegation summaries and identify whether Repeaters tend to follow pattern of behaviors when interacting with the victims of police brutality or complainant of the allegations.