Checkpoint 5 - Natural Language Processing

INTRODUCTION

Chicago PD offers rewards to their police officers to recognize their efforts on the job. Civilians file complaints against police officers based on their encounters. We compute a metric which we call ‘Complaint Severity Score’ or CSS based on the average annual complaints of an officer multiplied by custom defined weights based on the complaint category. The weights are defined based on research by our previous cohort and ranges between 0.0 - 43.0(here a higher score is subjectively worse). With our analysis on whether CSS is correlated with the awards police officers receive, we hope to understand whether ‘good’ police officers i.e. officers with low CSS scores are rewarded by the Chicago PD more than compared to cops with higher CSS scores.

MOTIVATION

For the natural language processing checkpoint, we want to find out whether there is any relationship between the CSS metric and the sentiment polarity of allegation summaries. If there is indeed a relationship between these two variables, then it tells us that CSS may be a strong and reliable metric by itself. This could be shown through a scatter plot or linear regression model. However, a cloudy formation on the scatter plot and / or a high p-value and R-square on the regression model summary would mean that these two variables are not related and potentially closes this avenue of exploring any association between allegation summaries and CSS by themselves.

ANALYSIS

We performed sentiment analysis using the Vader module from NLTK library, and we got output probabilities as positive, negative and neutral. In addition, we also got a sentiment polarity indicator called compound score, which ranges from -1 to 1. Where a score of -1 indicates a negative sentiment and +1 indicates a positive sentiment and 0 indicates a neutral sentiment. From the below scatter plot(fig 1.1) we can easily observe that there is absence of any strong correlation between CSS and Compound score, rather we are seeing a cloudy pattern.

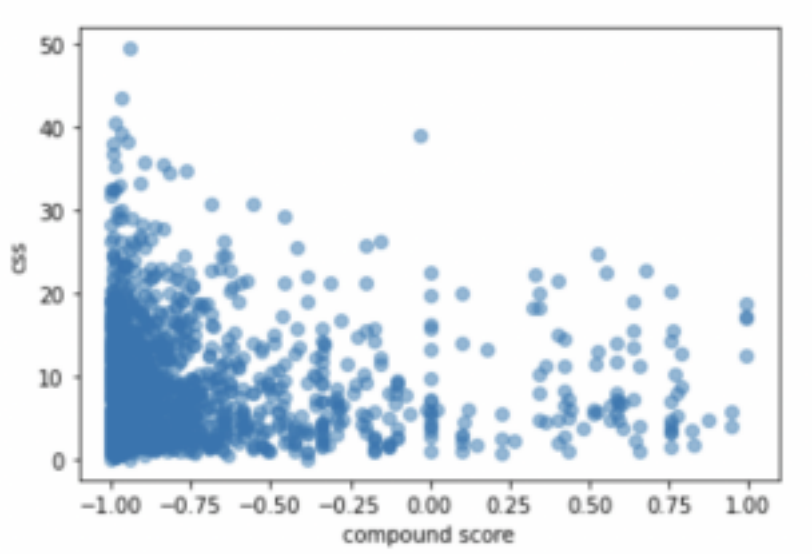

Fig 1.1

OLS Regression Results

Dep. Variable:	css	R-squared:	0.002
Model:	OLS	Adj. R-squared:	-0.000
Method:	Least Squares	F-statistic:	0.7910
Date:	Wed, 01 Dec 2021	Prob (F-statistic):	0.531
Time:	21:37:53	Log-Likelihood:	-5575.2
No. Observations:	1681	AIC:	1.116e+04
Df Residuals:	1676	BIC:	1.119e+04
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	328.5081	398.948	0.823	0.410	-453.980	1110.997
compound	0.4046	0.561	0.721	0.471	-0.696	1.506
pos	-324.5849	398.748	-0.814	0.416	-1106.681	457.511
neg	-315.7638	398.819	-0.792	0.429	-1097.999	466.472
neu	-320.0278	399.016	-0.802	0.423	-1102.649	462.594

Omnibus:	507.668	Durbin-Watson:	1.917
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1458.060
Skew:	1.555	Prob(JB):	0.00
Kurtosis:	6.339	Cond. No.	7.43e+03

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 7.43e+03. This might indicate that there are
strong multicollinearity or other numerical problems.

Fig 1.2

This observation is further supported by the linear regression model above(Fig 1.2) where we used compound score to predict the CSS metric. Looking at the F-Statistic score and P-value of compound, we can infer that there is no association between compound polarity score and the CSS of an officer.

CONCLUSION

In conclusion, we don't see any clear correlation between compound polarity score and the CSS of an officer. Since the data is so sparse(1681 rows) that the allegation summary may not be a good avenue to derive some insightful analysis using these variables alone.

Further Work (NLP)

MOTIVATION

Officers may potentially spend a lot of time writing a report and filling up an allegation summary form, when in fact their presence would be more valuable in the field fighting crime. To minimize their time writing such reports and filling up allegation summaries, we built an auto-predict algorithm using deep learning that will automatically predict and recommend the next 3 words for their report. This would allow them to finish up the writing faster, thereby minimizing the time needed on such administrative tasks.

ANALYSIS

We used a LSTM model to generate text a character-by-character. After running the model for 25 epochs, the model started generating locally coherent sounding texts. We ran the entire code on GPU, as the recurrent networks are quite computationally intensive. The data which we used is very sparse(1681 rows) and we would have achieved a much better performance if our corpus had at least ~100k characters. In the below figure(Fig 1.3) we tried generating the next three words based on the inputs of the user.

Diversity acts as a context regularization to basically show that how contextually relevant we want the auto generated texts to be based on the diversity score it generates the words. In our experiment, we found a diversity score of 0.2 to be a sweet spot where we are getting a very coherent sounding auto generated texts.

```
...Diversity: 0.2
...Input Sentence: " battery, and subsequently pleaded guilt"
...Autocomplete:  y of the complainant

...Diversity: 0.5
...Input Sentence: " battery, and subsequently pleaded guilt"
...Autocomplete:  y of the department
```

Fig 1.3

CONCLUSION

In conclusion, we think Chicago police can benefit heavily by using a system like this, which can help cut down the time it takes to file a report, which in turn will make them more productive.

Submitted by: The Brave Dingos (Keshav Bhandari,
Dhruv Narayan,
Akshay Kumar)