DSCI 510 Final Report:

Moral judgement and sentimental analysis on Chinese social media platform

1. Project description

The project investigates how moralized content influences user engagement and emotional expression on the Chinese social media platform. Using natural language processing techniques, the study extracts textual features and examines the interplay between these variables. Specifically, it measures moral and sentimental aspects of social media original posts and comments related to seven highly discussed events on Sinaweibo.com during October and November.

2. Data

2.1 Data collection

This project retrieved data from the dynamic Sinaweibo.com webpage using a scraper script programmed with Selenium and Chrome Driver. This method was chosen for two key reasons. Firstly, it enabled the extraction of real-time social media posts sorted by temporal proximity to the script running date, facilitating the detection of scraping failures due to the website's automated filtering algorithm. Secondly, unlike scraping under specific tags which only displays some relevant posts, this method allowed for a larger sample size.A total of 2337 posts were collected using this script. Subsequently, posts containing fewer than three Chinese characters were removed, resulting in 2107 source posts for subsequent analysis.

Table: Sample distribution

Keyword for search	Raw data count	Cleaned data count	Source posts count with comments data	Comments count
Openai	385	352	56	343
Artificial intelligence risk	153	148	27	87
Post deletion (content moderation)	386	350	137	812
Sexual assault	376	349	93	609
Homeless dogs hurt	299	280	76	634
Cat abusing	397	351	85	357
Du Meizhu scandal	341	277	45	117
Total	2337	2107	519	2959

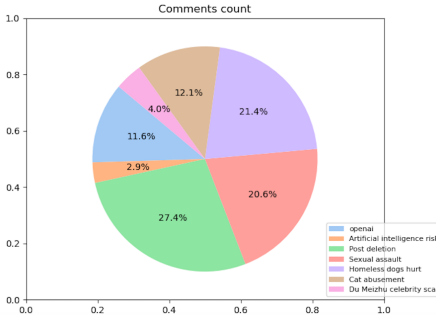
Following the extraction of source posts, a second scraper script was employed to gather comment posts by requesting data through the unique mid (16-digit identifier) of each source post. This method enabled the retrieval of all accessible comments, except those moderated by the platform algorithm. A total of 5081 raw comment data points were collected.

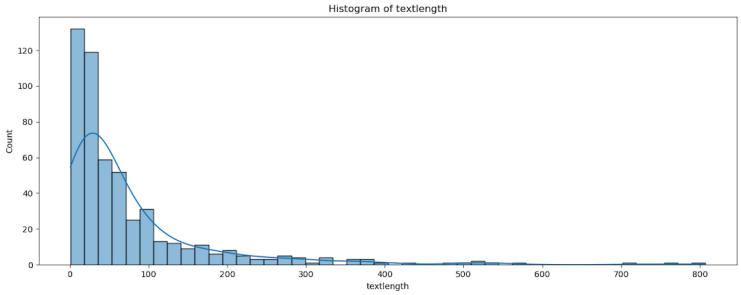
The raw comment data underwent cleaning based on three principles:
a) Removal of empty values.
b) Elimination of repetitive content and content moderation warnings displayed as "not accessible due to violation of platform regulation."
c) Exclusion of comments containing fewer than three Chinese characters, as these posts often lacked significant semantic content, comprising casual phrases like 'Good nights' or 'Here', which are unrelated to the original source post.

These three steps of screening tighten the comment dataset to 2959 samples, the distribution of which is shown in the *Comments Count* column. These comments are mapped to source posts by joining two datasets by column *mid*. The features of multiple comments from the same source post are averaged to represent the general comment style and sentiment of a source post.

2.2 Data descriptive analysis

The visualize_results.py script generates histograms, boxplots, and pie charts to illustrate the distribution of individual variables. A series of four pie charts presents the sample proportions across different search topics.





Examining histograms and boxplots for both post popularity metrics and textual features reveals the presence of high-value outliers. Therefore, numerical variables are grouped into categories or removed of outliers to ensure the robustness of further analyses. Boxplots and histograms are created after grouping textual features by popularity features (likes, reposts and comments count) and text length. Compared to popularity features, text length turned out to be a good dimension for grouping data.

Ultimately, a dataset comprising 519 source posts and 24 variables was constructed to examine the correlation between source posts and comment features. For detailed definitions and computation methods of all variables, please refer to the document *Final_report_APPENDIX_LeyiS.pdf*.

3. Analysis and Visualization

3.1 Analysis steps

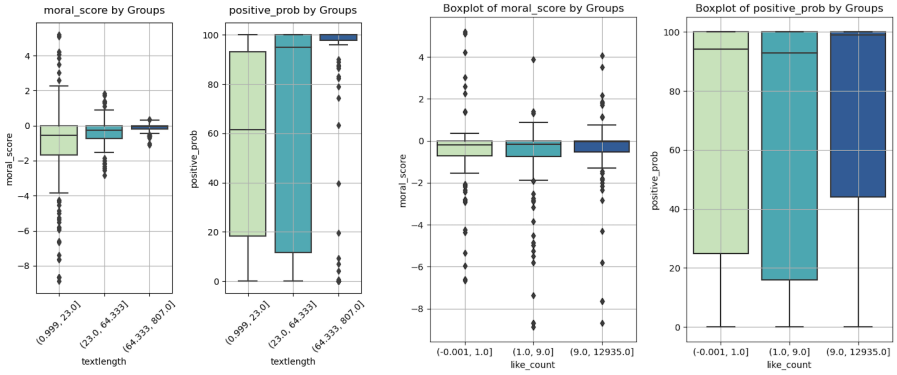
To explore the correlation between variables, I expanded the descriptive analysis by creating a heatmap and organizing data based on different variables to identify notable variations in metrics. Subsequently, I proceeded with statistical analysis, conducting one-

way ANOVA tests, and fitting linear regression models to uncover explanatory patterns within the datasets.

3.2 Findings

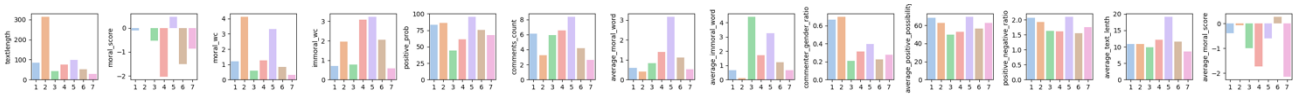
Finding 1: The distribution of posts' moral implications varies among different text length groups, while popularity measurements do not exhibit such distinctions. *(/results/correlation_plots)*

The segmentation into three groups via the q-cut method—texts under 23 characters, texts between 23 and 64 characters, and long texts exceeding 64 characters—reveals intriguing insights. Longer source posts tend to possess a moral score close to zero, displaying minimal variance, indicating a scarcity of moral judgment content. Conversely, shorter source posts are inclined to have immoral judgment words across a wide value spectrum. This observation suggests that longer posts on Sinaweibo.com potentially exhibit greater homogeneity, while shorter source posts depict a more diverse range of users' moral expressions. Consequently, the subsequent analysis adopts the proportion of moral judgment words as a more effective feature for ANOVA analysis.



Finding 2: Notable variations are evident in textual features among posts and comments originating from different events.

A set of visualizations was generated to explore these differences (Figure 1), comprising 13 bar charts, each representing a feature across seven selected events. The figure is plotted by aggregating total data based on the keyword column and computing average values for specific variables. Additionally, the same function was utilized to visualize median and sum values, offering a comparative view of inter-group variance *(/results/correlation).*



Subsequently, based on Figure 1, one-way ANOVA analysis was conducted on 16 variables to ascertain the statistical significance of variances between post topics. The analysis presented robust outcomes even after the removal of outliers that deviated more than 3 standard deviations from the mean. This analysis yielded following key insights:

a) **The presence of moral emotion words in source posts can influence commenters to post longer comments, but this effect appears to vary based on the post topic**. When not categorized by topics, factors such as sentiment positivity, count of moral

judgment words, and count of moral emotion words exhibit weak correlations with comment length (R^2 = 0.17, 0.22, 0.22). However, when grouped by topics, the analysis reveals interesting nuances. In the Du Meizhu scandal topic, a higher presence of moral emotion words in source posts correlates with longer user comments (R^2 = 0.44); while in the post deletion topic, more intense immoral judgments (mostly referencing social media platform over-moderation) correlates most strongly with longer user comments (R^2 = 0.42). The topics might have an intermediate effect that adjusts the correlation mechanism between textual features.

b) **Users express different moral emotions but share nonmoral feelings disregard of topics**, **while sentiment expressions vary between topics.** Regarding the sentiment expressions in comments, Openai, Ai risks, homeless dogs' management, and cat abusing have twice more positive comments than negative ones, while the other topics are balanced over the proportion of positive and negative comments. Regarding the sentiment of source posts, all topics except post deletion are given a high possibility score for expressing positive emotions. For moral emotion scores, the distribution within each topic is more scattered compared with nonmoral emotions, with many more outlier data points.

c) **Different genders prefer to leave comments on different topics.** There's a huge variance between the gender ratio of commenting users of each topic. For texts related to post deletion, sexual assault, cat abusing and Du Meizhu Scandal, the comments under each source post have an average of below 20% male users; while for posts from the topic of Openai and AI risk, an average of over 80% comments are left by male users. On one hand, this might characterize the different interest of gender groups. On the other hand, this also warns against the recommendation bias of social media algorithms.

Finding 3: Users' commenting behavior is likely to be motivated by personal moral judgements. *(/results/regression)*

Heatmaps and scatter plots are created to explore how moral judgements are expressed on Sinaweibo.com. Figure 3 depicts the proportion of moral and immoral judgement words within each source post has a negative correlation with the length of source post. The dependent variables are computed by dividing the moral /immoral word count by number of words in a post to control the collinearity of variables, considering that longer posts naturally have higher word counts for any type of word.

a) **Moral and immoral judgement in comments best predicts the number of comments one source post receives**. Though this could not be a cause-and-effect mechanism, the correlation between moral judgement words presence and the number of comments of a source post can be characterized well with a linear model; conversely, moral emotion or sentimental tendency from both source posts and comments do not show any significance. According to the linear model summary, an average increase of 1 moral judgement words of each comment under a source post corresponds with an increase of 11 more comments being made.

b) **The moral emotions and immoral judgements go together in comments.** The linear model is statistically significant, and one point of increase in moral emotion score comes with 10 more immoral judgement words from source post comments. However, nonmoral emotions and positive/negative sentiment tendencies do not show significant linear relationship with commenter's moral judgements.

3.3 Impact and conclusions summary

The above analysis raises attention towards specific variables that might have a strong influence on analyzing social media textual features. Past literatures seldom take post topic into consideration due to the challenge of collecting large datasets and immature sampling principles. This test on small datasets reveals that detecting the object or topic of sentimental or moral expressions is of great importance.

Due to the limit of datasets quantity in this project, main takeaways from the analysis point to the reflection on computational tools and models.

Firstly, topic difference is not measured by the current general moral behavior and judgement dictionary. Semantic features of social media post topics have significant effect on how moral emotions are expressed and how users make moral judgements in reaction. Topic computation should be considered and incorporated into moral computation and value computation.

Secondly, the project questions the results from existent studies focusing on western social media context and limited case studies. Moral scores and sentiment scores are weakly correlated with most variables, which are not aligned with past literature. This result calls for further tests on the efficacy of feature computation methods. For example, when using moral behavior dictionary and sentimental classification models to one dataset, it's important to check whether moral judgement words have inner sentimental inclination which may affect feature extraction from all aspects.

4. Future work

Future work will be continued to apply the moral behavior dictionary and moral emotion computation on larger datasets. I plan to use opensource datasets and involve manual annotation on sub samples to test the efficacy of the tools. Other feature extracting methods will be tested together with the dictionary-based method to compare the performance of models and find the best measurement for people's expressed morality. For example, applying word embedding models on sentence level instead of the moral score computed by graph centrality, and substitute current general sentimental analysis model with a new one fine-tuned on social media data.