Exploring Online Adoption Discourse in r/Adoption and r/Adopted

Ethan Kozlowski

SOCI 41033 Computational Content Analysis

March 8, 2024

James Evans

Word Count:

INTRODUCTION

Adoption is a global industry and affects the lives of millions of people in the world. As a means of identity construction and validation, individuals explore online communities

RESEARCH QUESTIONS

How do the conversational dynamics and post sentiments within online adoption communities differ between general-purpose adoption subreddits like r/Adoption and adoptee-exclusive spaces like r/Adopted? Do adoptees and non-adoptees differ in their sentiment towards adoption within these different spaces? To help answer this main research question, I will also propose a handful of supplementary research questions. Are the topics discussed in r/Adoption and r/Adopted distinctly different from one another? I have a hunch that even though both subreddits talk were formed to discuss issues related to adoption, in reality the individuals populating these spaces discuss various aspects of it. Since more adopters post in r/Adoption, there are more discussions about the process of adopting. Whereas in r/Adopted there might be more discussion about trauma. Are the posts from adoptees distinctly different from posts not made by adoptees in r/Adopted and r/Adoption? If the answer is no, then the main question becomes much harder to answer. This is a classification question. Because not all users say whether they are adopted or not, it is not easy to understand who is and is not an adoptee when looking at these conversations.

DATA COLLECTION

I gathered data from two main sources. Firstly, a large portion of the posts and comments come from a Reddit archival site. r/Adoption is a sufficiently large subreddit, and was, therefore, archived on this site. One issue with this data was it was not up to date past December of 2021. Additionally, my other subreddit of interest—r/Adopted—was too small of a community to be picked up by this site to be archived. Therefore, I developed a customized web scrapper to help me find Reddit posts from old.reddit.com, a simplified version of reddit.com that did not require the use of dynamic scraping packages like selenium. A decision was made early on in this paper to scrape this non-dynamic version of Reddit due to issues loading the selenium package for dynamic scraping; despite trouble shooting this issue for a few weeks, I made the ultimate decision to go ahead with the somewhat flawed, but still useful data collection methods I had at my disposal.

ANALYSES

*Machine Learning*

An important task to aid in the analysis of this project is to answer this supplementary research question: Are the posts from adoptees distinctly different from posts not made by adoptees in r/Adopted and r/Adoption? Fundamentally, this involves a supervised Machine Learning classification task. As a result, I developed multiple ML models to attempt to classify the title combined with the post text of reddit threads into a binary classification: adoptee or non-adoptee.

*Word Embedding*

*Topic Modeling*

*Sentiment Analysis*

According to Hutto and Gilbert (2014), VADER (or Valance Aware Dictionary for sEntiment Reasoning) is a powerful tool for sentiment analysis, generally outperforming individual human raters. Because VADER is meant for use on social media data, I found the model particularly enticing. I incorporate this sentiment analysis tool to gather the compound sentiment from my cleaned but not tokenized data. Though I had originally considered training a fine-tuned LLM like Chat GPT to conduct a sentiment analysis, the costs of running these analyses seemed prohibitive and I instead opted for the simpler use of VADER. After conducting a 2-way analysis of variance test, the results of the statistical test are shown in the table below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **sum\_sq** | **Degrees of Freedom** | **F-stat** | **PR(>F)** |
| is\_adoptee | 13.540146 | 1.0 | 36.756788 | 1.350233e-09 |
| subreddit | 121.939295 | 1.0 | 331.022779 | 1.144487e-73 |
| is\_adoptee \* subreddit | 0.000014 | 1.0 | 0.000038 | 9.950871e-01 |
| Residual | 14685.122317 | 39865.0 |  |  |

DISCUSSION

CONCLUSION

*Improvements and Future Directions*

I recognize that this research paper has a few faults. One remedy for future works would be to adjust the intake of posts from both subreddits to be more equal. r/Adoption is a much larger and more active subreddit than r/Adopted, therefore I should have under sampled from this subreddit and oversampled from r/Adopted to keep the data close to being evenly distributed. Furthermore,

There are many possible avenues for future research.

Works Cited/Consulted

Chen, Yea-Wen and Masato Nakazawa. “Emotions and Pan-Asian Organizing in the U.S. Southwest: Analyzing Interview Discourses via Sentiment Analysis.” *Voluntas,* 28:2784-2806. doi:10.1007/s11266-017-9916-3.

Gentzkow, Matthew and Jesse. M Shapiro. 2010. "What Drives Media Slant? Evidence from U.S. Daily Newspapers." *Econometrica,* 78(1):35-71. Retrieved January 13, 2024, (https://www.jstor.org/stable/25621396).

Hutto, C.J. and Eric Gilbert. 2014. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Association for the Advancement of Artificial Intelligence*. Retrieved March 8, 2024, (https://ojs.aaai.org/index.php/ICWSM/article/download/14550/14399/18068)

Davis, Jenny L., Tony P. Love, and Phoenicia Fares. 2019. “Collective Social Identity: Synthesizing Identity Theory and Social Identity Theory Using Digital Data.” *Social Psychology Quarterly,* 82(3):254-273. doi:10.1177/0190272519851025.

Stuhler, Oscar. 2021. "What’s in a category? A new approach to Discourse Role Analysis." *Poetics,* 88. doi:10.1016/j.poetic.2021.101568.

Waller, Isaac and Ashton Anderson. 2021. “Quantifying social organization and political polarization in online platforms.” *Nature,* 600(7888):264-268. doi:10.1038/s41586-021-04167-x

Yadlin-Segal, Aya. 2020. “Articulating Persian identities between Iran and Israel: On nationality, diasporas, and lived ethnicities in online media.” *Nations and Nationalism,* 27:347-362. doi:10.1111/nana.12639.