Schizo-Media: Quantifying Sentiment in Social Media Interactions with Schizophrenics

Joshua Munsey

Reedley College

Honors-3A

Dr. John Heathcote

29 April 2021

Abstract

The difference in written language on social media of those with and without schizophrenia online was quantified in a previous study (Mitchell et al., 2015). Given this information, we have been able to further explore how each group's social media peers respond to said differences. By contrasting the sentiment of responses received by those with schizophrenia versus a control group, we have gained key insight into whether there exist potential risks (e.g., cyberbullying) in those with schizophrenia using social media as a therapeutic means (Naslund et al., 2016). Furthermore, such research has been able to evaluate the treatment of people with schizophrenic symptoms within normal online discourse. Attached with a term frequency dictionary and an LDA clustering method, we have been given further insight into what differences do exist and if parts of the language used may be the cause of said differences.

Introduction

Schizophrenia is a psychotic mental disorder in which roughly 1% of the population is afflicted. Those living with this disorder experience symptoms such as hallucinations, delusions, and disorganized or catatonic behaviors. Further research has also unveiled that this illness is responsible for compromising much of the social functioning in those who are afflicted (Aboitiz & Billeke, 2013).

Given the debilitating nature of such symptoms, there has been a subsequent need for exploration in potential therapies to assist in the socialization of those with schizophrenia. One such suggestion was the leveraging of social media as a form of peer-to-peer support system, where those who have schizophrenia or schizophrenia adjacent disorders gain the opportunity to champion their disorders or gain insight into how others afflicted with similar disorders live their daily lives (Naslund et al., 2016). While there may be some conflict on how social media affects the mental health of their users (Abdalla et al., 2020), this avenue of therapy shows potential short-term improvement in assisting those with schizophrenia to socialize.

With previous research being done in how those with schizophrenia interact in the format of social media (Mitchell et al., 2015), it is now paramount to evaluate the perceived risks (e.g., harassment) associated with social media playing a supportive role in their lives and how their experiences may differ. This research aims to accomplish just that through the evaluation and quantification of sentiment in the responses received by schizophrenic users. Specifically, this research has entirely focused on responses received to public posts on the social media platform Twitter. A public post is treated as an open forum where stranger and friend alike are given an opportunity to respond, ultimately expanding the lens in which we can view the conversation of social media and schizophrenia while drawing comparisons to a control group for reference.

Methodology

This research was primarily carried out with the utilization of a linear Support Vector Machine (SVM) to quantify sentiment of responses received by those with schizophrenia compared to that of a control group. A Support Vector Machine is a type of machine learning algorithm that encodes human annotated data into numerical matrices based upon features of said data, finds a hyperplane between the known classes of data, and subsequentially repeats the process, classifying any new unannotated data based upon which side of the hyperplane it falls. Following the implementation laid out by Cambria et al. (2015), we have implemented a promising multi-feature and multi-class Support Vector Machine. However, given the nature of our data in it being responses, we modified the suggested feature list, removing the "@USER" feature encoder, giving our Support Vector Machine a forecasted accuracy shown in Figure 1. Our classes similarly included Positive, Neutral, and Negative sentiment. For training data, we utilized a combination of the Sentiment140 and the Twitter Airline Data's neutral sentiment dataset. The accuracy score we received for the Support Vector Machine is shown below in Figure 2.

After initial results provided by the above Support Vector Machine were reached, we accomplished further contextualization of our data through clustering and term frequency counts. The former involved the utilization of LDA and Brown clustering, LDA clustering being responsible for pulling topics out of the text and Brown clustering pulling semantic usage of words used. Term frequency counts are merely a count of how many times a word appears excluding stop-words (e.g., "the" and "a").

Finally, for data acquisition, we utilized the same standard for finding schizophrenic users as used by Mitchell et al. (2015). This method being the self-admission of being diagnosed with schizophrenia on a public tweet. We further leveraged the Twitter API to collect responses from any further public posts of users with this prior self-admission. This data was collected from users within the five-year gap 2016 and 2021. Similarly, the control group users were collected randomly between that five-year gap as well and cross referenced to ensure no single user appeared in both groups before collecting an equal number of responses from their public posts.

Features	Positive			Negative			Neutral			F_{pn}
	P	R	F	P	R	F	P	R	F	
All Features	0.824	0.629	0.713	0.612	0.607	0.610	0.679	0.831	0.748	0.662
w/o N-grams	0.671	0.597	0.632	0.430	0.574	0.491	0.645	0.637	0.641	0.562
w/o POS Tags	0.814	0.611	0.698	0.633	0.589	0.610	0.669	0.839	0.744	0.654
w/o @User, Hashtag, URL, Discourse	0.821	0.616	0.704	0.602	0.607	0.605	0.672	0.826	0.741	0.654
w/o Sentiment140	0.814	0.616	0.701	0.602	0.599	0.600	0.676	0.830	0.745	0.651
w/o Bing Liu	0.821	0.621	0.707	0.616	0.603	0.610	0.676	0.833	0.746	0.658
w/o NRC Emotion + Hashtag	0.816	0.619	0.705	0.609	0.597	0.603	0.676	0.832	0.746	0.654
w/o SentiWordNet	0.821	0.624	0.709	0.610	0.597	0.603	0.674	0.830	0.744	0.656
w/o SenticNet	0.820	0.615	0.703	0.610	0.597	0.603	0.674	0.837	0.747	0.653
w/o Negation	0.811	0.610	0.701	0.598	0.601	0.593	0.674	0.824	0.744	0.647

Table 1. Forecasted accuracy given our modifications according to the original SeNTU paper. With each number representing the Precision, Recall, and Accuracy with 1.0 representing 100% in that category.

	precision	recall	f-score
Negative	71%	67%	69%
Neutral	89%	95%	92%
Positive	70%	70%	70%

Overall Accuracy: 77%

Table 2. Accuracy of Support Vector Machine we achieved after training.

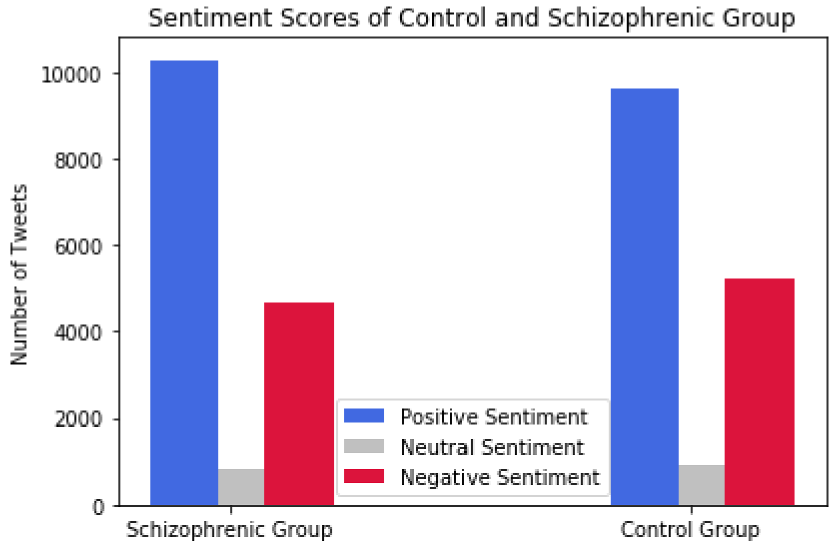
Precision being a ratio of True Positive over True Positive + False Positive.

Recall being a ratio of True Positive over True Positive + False Negative.

And F-score being 2 * Recall * Precision over Precision + Recall.

Results

Our Support Vector Machine ultimately found an inconsequential difference within the overall amount of positive, neutral, and negative responses compared to the control group. As can be seen in Figure 1, there is an insignificant difference with slightly more negative responses towards control users and low counts of neutral responses to both.




```
Positive Responses Schizophrenic Group: 10288
Neutral Responses Schizophrenic Group: 841
Negative Responses Schizophrenic Group: 4650

Positive Responses Control Group: 9631
Neutral Responses Control Group: 917
Negative Responses Control Group: 5231
```

Figure 1. The amount of positive, neutral, and negative responses each group received to public posts.

Further similarities appear within the term frequencies as well with only slight variations as shown in Table 3.

word	schizo	control
like	820	574
get	567	467
I'm	526	300
people	467	211
know	466	358
would	448	229
one	446	344
Thank	423	204
think	421	256
good	405	267

Table 3. Most frequent words with their respective number of appearances.

Similarly, the Brown clustering found little information not provided by the previous term frequency count. However, the LDA clustering method we employed found slight variations in topics discussed between each group. Such topics appearing to be most prevalent within the schizophrenia group involve “god”, “support”, “love”, and “talk”. In contrast, the most prevalent topics within the control group appear to include the words “Phone”, “call”, “hope”, and “damn”.

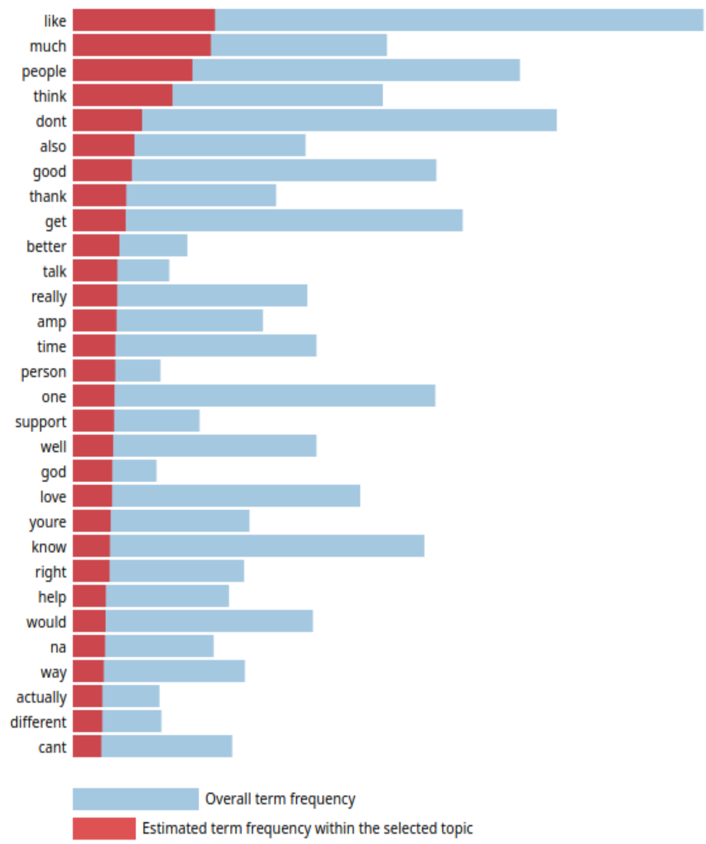


Figure 2. Most frequent words within most populated topic for the schizophrenic group.

Discussion

Social Media continues to be a relatively new and rapidly changing tool that requires much further investigation with many risks still to be explored before consulting those with mental disorders to seek support from it. However, from the results we have received, we can conclude that those with specifically schizophrenia are treated relatively similar to those without, regardless of any differences in textual language usage on the platform. This could suggest indifference to the language difference or merely that the differences may not be noticeable. Furthermore, the LDA cluster seems to suggest that there may be a pattern of support towards those with schizophrenia in line with the proposal put forth by Naslund et al. (2016). Ultimately, this added context introduced into the conversation of social media and its role as a supportive outlet for those with schizophrenia can further help in discussing the feasibility of different therapies to assist in socialization.

References

Abdalla, Lamis F. et al. "Social Media Use and Its Connections to Mental Health: A Systematic Review." *Cureus, 12*(6), e8627. https://doi.org/10.7759/cureus.8627

Aboitiz, Francisco & Billeke, Pablo. "Social cognition in schizophrenia: from social stimuli processing to social engagement." *Frontiers in psychiatry, 4*, 4. https://doi.org/10.3389/fpsyt.2013.00004

Prerna, Chikersal et al. "SeNTU: Sentiment Analysis of Tweets by Combining a Rule-based Classifier with Supervised Learning." *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*.

Mitchell, Margaret., et al. "Quantifying the Language of Schizophrenia in Social Media." *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*.

Naslund, J. A., et al. "The Future of Mental Health Care: Peer-to-Peer Support and Social Media." *Epidemiology and Psychiatric Sciences*, vol. 25, no. 2, 2016, pp. 113–122., doi:10.1017/S2045796015001067.