Negative vaccine voices in Swedish social media

Dimitrios Kokkinakis[1], Mia-Marie Hammarlin[2]
[1]Department of Swedish, Multilingualism, Language Technology, University of Gothenburg, Sweden
[2]Department of Communication and Media, University of Lund, Sweden

https://doi.org/10.36505/ExLing-2022/13/0025/000567

Abstract

Vaccinations are one of the most significant interventions to public health, but vaccine hesitancy creates concerns for a portion of the population in many countries, including Sweden. Since discussions on vaccine hesitancy are often taken on social networking sites, data from Swedish social media are used to study and quantify the sentiment among the discussants on the *vaccination-or-not* topic during phases of the COVID-19 pandemic. Out of all the posts analyzed a majority showed a stronger negative sentiment, prevailing throughout the whole of the examined period, with some spikes or jumps due to the occurrence of certain vaccine-related events distinguishable in the results. Sentiment analysis can be a valuable tool to track public opinions regarding the use, efficacy, safety, and importance of vaccination.

Keywords: vaccine hesitancy, sentiment analysis, Swedish, exploratory study

Introduction and background

According to the World Health Organization (2019) vaccine hesitancy, "the reluctance or refusal to vaccinate despite the availability of vaccines", was one of the top-10 threats to global health even before the pandemic. Although vaccinations are considered as one of *the* most significant interventions to public health, vaccine hesitancy and resistance creates serious concerns for a significant part of a population in many countries, including Sweden. Vaccine discussions are often taken online in various social media platforms. For many people, the use of such platforms is the major source for information related to health issues including vaccinations (Klimiuk et al., 2021). Consequently, the availability of such digital content enables researchers to rapidly analyze and monitor large amounts of data, to e.g., identify and better understand the vaccine-deniers' arguments against vaccinations which in turn, can rapidly be spread as *rumours* to an even wider audience. The aim of this study is to investigate the Swedish public sentiment reflected in discussions related to the COVID-19 vaccines (Scannell et al., 2021; Yousefinaghani et al., 2021). Sentiment analysis could guide us to identify, classify and quantify posts related to vaccine hesitancy and provide useful information to assist public health experts and pro-vaccine organizations to formulate even more targeted policies and strategies to reduce anti-vaccine reactions. This work is part of a project

that investigates vaccination skepticism, and how it can be understood as an expression of civic engagement in the present digital times.

Data collection and preparation

The data in the study comes from a popular Swedish open discussion platform *Flashback Forum*. Fourteen threads have been chosen that discuss topics related to vaccination, such as *Vad ska man göra om man vägrar ta corona-vaccin?* 'What to do if you refuse to take the corona vaccine?'; collected prospectively between Sept. 2020 – Jan. 2022; 21,000 unique posts from over 2000 unique users. For each post, we stored the text and relevant metadata such as the date of publication and the number of posts produced by each user. During preprocessing, duplicate posts were removed and transformed to lower case. The dataset was further tokenized (separating punctuation and metadata from words), while emoticons were converted to a text description (😎: <smiling face with sunglasses>). Multiword expressions and phrasal verbs were also recognized, and their contiguous components were joined with an underscore prior to further processing (*spanska sjukan* 'the Spanish flu'; *spruta in* 'to inject').

Sentiment analysis and sentiment shifters

Sentiment analysis (Liu, 2012) is an application that infers subjective aspects of writing such as opinions, feelings, and classify the polarity of a text as *positive*, *negative*, or *neutral*. Sentiment analysis is a technology that allows us to gain a wider public opinion behind certain topics and a picture of how public opinion develops over time. There are many ways to approach the task (Birjali et al., 2021). Here, we apply a "classical" dictionary-based approach having a Swedish word list as its main component (*SenSaldo*; Rouces et al., 2018). Each tokenized text fragment in each post, is assigned a sentiment value according to *SenSaldo*. Moreover, since the polarity of individual words can change, when these are used in specific contexts, we also identify *sentiment shifters*, i.e., modifiers that change the sentiment orientation of words, e.g., via negation. Finally, the sum of the words' sentiment values for each post are summarized and scored, as the ratio of words' sentiment values by the total number of tokens in a post. E.g., for the fragment: *jo det är ni idioter som är hysteriska och de verkliga träskallarna* 'well it's you idiots who are hysterical and the real numbskulls' there are 3 underlined negative words found out of the 12, which returns a negative score of –25.

Exploratory analysis

Sentiment analysis is the practice of extrapolating the sentiment of an idea or event by automatically classifying written texts as some polarity value. In the specific context of *vaccine hesitancy/skepticism*, the people's attitudes present a dominant negative polarity throughout the whole covered period with 56.2% negative, 25.1% positive and 18.7% neutral posts. Figure 1 (left) shows the 25 most frequent positive (left; the most frequent is: 'fullt frisk' *completely healthy*) and negative (right; the most frequent is: 'allvarlig sjukdom' *serious disease*)

multiwords in the data as word clouds. Figure 1 (right) shows the proportion of positive/negative users in thread t3342731: *Vaccinerade, när kan ni tänka er ta en tredje dos?* 'Vaccinated, when can you imagine taking a third dose?'. Using sentiment analysis, negative voices can be identified and further analyzed using *qualitatively driven mixed methods* such as the sociologically-grounded narrative typology (Frank, 2013), and the exploration of the civic aspects of ethnological research. These are research directions that we have already started to explore in this project (cf. the Acknowledgements).

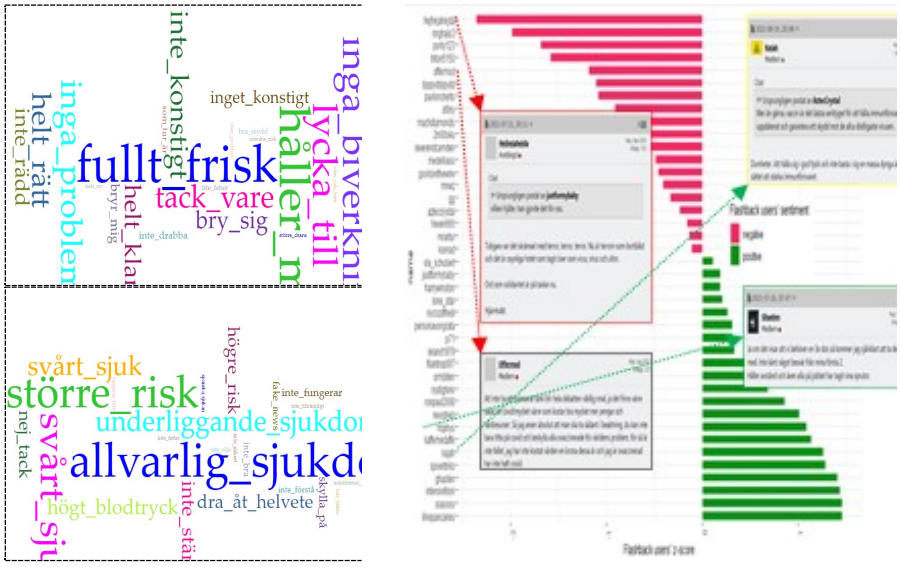

Figure 1 (left). The 25 most frequent *multiword* sentiments in the whole data (positive at the top and negative at the bottom). Figure 1 (right). The proportion of positive/negative users in thread t3396531.

Conclusions, limitations and future work

It is well established that the Internet provides valuable resources to fight vaccine hesitancy, providing insights to vaccine-undecided individuals concerns. Digital technologies such as sentiment analysis are important tools which can be used for gaining helpful insights into various complex health related topics (cf. Sandhya et al., 2021). Since this is an exploratory study there are many questions left for future work, for instance to investigate whether the positive sentiment messages are the pro-vaccination messages while the negative sentiment messages are the voice of the opposite. The key motivation of sentiment analysis is to analyze a body of text for understanding the opinion expressed by it. Public sentiment is important to e.g., determining appropriate interventions and policies, and therefore these techniques have been used in

many scientific and commercial applications. Nevertheless, sentiment analysis has its limitations with respect to accuracy, which can be lower for certain types of creative language use such as sarcasm and irony (Hernández Farias & Rosso, 2017). In the future, we would like to improve the accuracy of the sentiment analysis tool by employing more advanced techniques that can also capture subtleties of opinion and emotions beyond simple polarity (Izard, 2009). The dictionary-based approach, perhaps is not the best technique (Atteveldt et al., 2021). Thus, to demonstrate the effectiveness of this approach, a comparison between different methods (e.g., dictionary *vs* machine learning) is left for future research. Another limitation of this work is the use of a small set of social media posts, and from only one available site. This might cause a sampling bias, therefore, soon we plan to enhance the dataset with posts from two other major social media sites, such as *Familjeliv* (www.familjeliv.se/).

Acknowledgements

This work is part of the project *Rumour Mining* (MXM19-1161:1) financed by the *Bank of Sweden Tercentenary Foundation* (Riksbankens Jubileumsfond).

References

Atteveldt, et al. 2021. The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches & ML. Com Meth Meas. 15:2

Birjali, M., Kasri, M. and Beni-Hssane, A. 2021. A comprehensive survey on sentiment analysis: Approaches, challenges and trends. Knowledge-Based Syst. 226.

Frank, AW. 2013. The wounded storyteller: body, illness, and ethics. 2nd ed., UCP.

Hernández Farias, DK and Rosso P. 2017. Irony, Sarcasm, and Sentiment Analysis. In Sentiment Analysis in Social Networks. Pozzi F. (et al.) (eds). 113-128. M Kaufm.

Izard, CE. 2009. Emotion theory and research. Ann Rev Psych. 60:1-25.

Klimiuk, K., et al. 2021. Vacc. misinf. on social media–topic-based content & sentiment analysis of Polish vaccine-deniers' comments on FB. Hum Vacc & Immunoth 17:7.

Liu, B. 2012. Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers.

Rouces, J., Tahmasebi, N., Borin, L., and Rødven Eide, S. 2018. SenSALDO: Creating a Sentiment Lexicon for Swedish. 11th Lang Res and Eval (LREC). 4192-98. Japan.

Sandhya, A., et al. 2021. Information Extraction & Sentiment Analysis to gain insight into the COVID-19 crisis. In Advances in Intelligent Syst & Comp 1387. Springer.

Scannell, D., et al. 2021. COVID-19 Vaccine Discourse on Twitter: A Content Analysis of Persuasion Techniques, Sentiment and Mis/Disinformation. J Health Com. 26:7.

Yousefinaghani, S., et al. 2021. An analysis of COVID-19 vaccine sentiments and opinions on Twitter. J of Inf Dis. 108, 256-262.