

Author: Jacob Fuehne

Date: 10/26/2020

Research Paper Summary

SOLO: A Corpus of Tweets for Examining the State of Being Alone, *LREC*, 2020

1. Problem definition and the main ideas of the research

The state of being alone can have a strong impact on people's lives, although the experience from time alone varies significantly among people. To distinguish between the different ways being alone is experienced, the terms loneliness and solitude are established. Solitude being the positive state of voluntary aloneness, and loneliness being the negative state of dissatisfaction with the quality of one's social life. In this paper, an annotated corpus called SOLO (State of being Alone) is created to provide data for the different states of aloneness through the query terms *solitude*, *lonely*, and *loneliness*.

2. Significance of research study (Importance and Challenges of research problem)

The SOLO corpus stands to help researchers in identifying trends in communication surrounding the different expressions of being alone. Because the feelings of loneliness are so closely related to depression and suicidal ideation, this research can also be beneficial in improving systems for NLP assisted mental health interventions and AI therapists programs. What makes this work challenging is that in psychology, the theories surrounding what causes solitude vs loneliness in people and how emotions work are not definitive and there are still many unknowns.

3. Main research questions and assumptions

This research seeks to answer several questions posed in the introduction of the paper surrounding how people use solitude, alone, and loneliness, comparisons between solitude text

and loneliness text, associated words with the queries, and determining demographic differences between perceptions of the state of being alone. These research questions are sought through analysis of the corpus. In the annotation and gathering of the corpus, it is assumed that by manually combing through search queries for words relating to the state of being alone for the most consistent results will give a dataset of tweets that is large and generalizable.

4. Research Methodology

The data is constructed by manually narrowing down a list of possible search queries to just solitude, lonely, and alone, as these were found to be the most consistently associated with the state of being alone. Using the Twitter API, 4 million tweets are gathered that contain these tokens. In this paper, no computational models are tested, however, the NRC Valence, Arousal, and Dominance lexicon and NRC Emotion lexicon are used as abstract models to determine the emotion associations of the words in SOLO.

5. Experiments

The newly created SOLO corpus is analyzed for various questions. The word solitude is found to more often describe positive self-imposed states of being alone, while lonely and loneliness are for negative and undesirable experiences, with loneliness being found in more negative contexts than lonely. It was found that women tweeted 12% more than men about being lonely, despite men making up 10% more of the data. Adolescents were also shown to be more likely to be associated with the negative experiences of being alone than adults. Overall the results serve to roughly verify previous theories on how certain demographics experience being alone.

6. Discussion

6.1 Important aspects

- Making a large, unique corpus publicly available

The SOLO corpus serves to help computational linguists and psychologists in working on research and methods related to how people experience being alone. This builds a good foundation for future researchers to build on in this area.

- Working on an important and neglected area for a corpus

Due to the issues of depression and suicide ideation often associated with being alone, this research is especially important in handling this mental health issue. As previously mentioned, this public corpus will reduce the barrier to entry and encourage for researchers to work on handling this problem.

6.2 Limitations of the paper

- Corpus was not tested with any computation models

While the paper stands on its own achievements, it would have been beneficial if there had been some baseline performances established when using this corpora, as had been done in the paper on complaints that we previously discussed this semester.

6.3 Questions for presenter

- If you were implementing a program using the SOLO corpus, do you think that you would simply combine the tags for lonely and loneliness into the general state of negative aloneness and compare that with solitude, or are lonely and loneliness distinct enough in this corpus to warrant keeping them separate?