

Neuro Linguistic Programming

Mikkel Werling
Victor Møller Poulsen

December 2020

1 Abstract

2 Introduction

2.1 From qualitative survey to quantitative data

The project is building on the work by Marco et. al (2020), which investigated ...

2.2 Introduction to the hypothesis

On the basis of the paper by Marco et. al (2020), we investigated two hypotheses...

3 Methods

3.1 Topic Modelling

Using LDA, underlying assumptions and hyperparameters.

3.1.1 Different Metrics to Evaluate Quality of Topics

Coherence as well as Cao Juan and Arun to get different measures of quality. We also looked at interpretability as...

3.1.2 Different atomic units for creating documents

3.2 Thoughts on choosing "atomic units"

Tree was higher in coherence, while thread was higher in the measure from Arun. So we had to take a more qualitative dive. Here we found that tree had several categories that seemed random (3 random words categories). Thread had overlap across categories, but seemed to capture more specific conversations. These were mainly due to different companies that people used. Moreover, the trees were made by pasting the submission text, which gives the submission text a lot of weight in the topic models. Here we focus on the common denominators between the two.

3.3 Topics and their distributions

We followed the division of the topics proposed in the qualitative study. Overall, we find that comfort and lighting as well as control and connectivity are the central themes of posts on the subreddits.

4 Results

4.1 H1

Do we find the same topics as the qualitative study?

- We find topics that resemble those found in the qualitative study to a large degree
- Light, connectivity and comfort dominates the topics
- First indications that most topics are concerned with "how to" and not a discussion board

4.2 H2

Are people even remotely concerned about privacy, security and trust? No.

- Trust and privacy are very rarely used on the forums (insert number)
- Security is the only word of interest which is used regularly. But it is used in combination with security systems and security cameras, not concerning the security of my data.

5 Discussion

5.1 Reddit as a "how-to" site

From the topics extracted from the corpus, it seems like the subreddits are primarily used to ask questions concerning specific products and how to implement them. Importantly for H2, it is used to much less degree as a discussion forum (for instance regarding security).

5.2 Use of seeded topic models?

We initially thought that seeded topic models would help us disentangle different senses of security. But it does not seem to be warranted - we do not find indications that security is used in the more abstract sense. However, security does come up in several

different categories. Notably, it comes up in topics related to cameras, which could indicate worries about cameras surveying them or just interest in security cameras. To investigate this, we sample some of the most representative documents of the topics and investigate the tendency. NOTE: Right now this is placed in the bottom of "doc_from_topic.py", but could be in its own document. But the findings are very clear - people are worried about buying the right camera, not whether it is filming them.