

Review of Master's Thesis by Enlik "Topic Modeling for Requirements Engineering : An Analysis of RideSharing App Reviews"

The thesis uses topic modeling techniques to identify the topics discussed in user reviews of ride-sharing apps. The method is validated on Europe's top 10 ridesharing apps. The contribution of this thesis can help the developers of ride-sharing apps in prioritizing their efforts when improving the quality of their apps.

The first main goal of the thesis is to discover latent topics from a collection of use reviews belonging to ride-sharing apps. Before running topic models (i.e., LDA standard, LDA Mallet, and BERTopic), the author has performed the following preprocessing steps on the review data: 1) removing non-English reviews, 2) removing inconsistent reviews, and 3) removing uninformative reviews.

The thesis has used a coherence score metric to find the optimal number of topics. However, the details about how the coherence score is calculated are not included in this thesis. Moreover, there is no justification provided as to why only the top 4 keywords per topic are presented when the BERTopic model is used. The list of keywords (under the same topic) extracted through BERTopic (i.e., prices and price, ride and rides, enjoy and enjoyed) clearly shows that lemmatization or stemming is not performed before applying BERTopic. Besides that, some stopwords such as *this*, *but*, *my*, and *he* are also detected as topic words. However, there is no discussion about those results in the thesis.

The second part of the thesis presents the analysis that shows the proportion of topics in 10 ridesharing apps and the average rating and sentiment score against each topic in ridesharing apps. Such an analysis can help app developers to identify the topics that should be the focus of the next development cycle of their app.

Overall, the form of the work is acceptable: the language is understandable, the structure is reasonable, and most references are relevant. Altogether the work satisfies the requirements for a master's thesis.

The grades for the different aspects of the work:

Content: 3.5

Complexity: 4

Form: 4

Questions to the author:

1. A coherence score metric is used in this thesis to choose the optimal number of topics. Why the plot in figure 8 don't show the coherence score versus the number of topics separately for each topic modeling technique?
2. Although inconsistent reviews have been removed in the preprocessing step, why the average rating and sentiment scores for some topics are in conflict with each other? For instance, for the topic 'pricing' in the case of BlaBlaCar, the average sentiment score is -1.1. However, the average rating score is 4.0
3. There are many off-the-shelf sentiment analysis tools available in the market such as Stanford CoreNLP, and NLTK/VADER, some are even tailored for the software

engineering domain (i.e., SentiStrength-SE, Senti4SD, and SentiCR) what was the reason behind selecting the tool SentiStrength over other sentiment analysis tools?

4. What practical issues do you see if the approach proposed in this thesis is used by ridesharing companies to understand users' requirements?

Faiz Ali Shah, PhD

Lecturer of Software Engineering

Institute of Computer Science at University of Tartu

17-08-2022