CERIS : Civil Engineering Research and Innovation for Sustainability

Automated text analysis on open-ended response surveys: measuring attitudes regarding autonomous vehicles

PhD student
Vishnu Baburajan

PhD program
Transportation Systems (IST, University of Lisbon)

Supervisors
João Abreu e Silva (CERIS, IST, University of Lisbon) and Francisco Pereira (Technical University of Denmark)

Co-supervisor
-

Period
2016-2021

Funding
FCT scholarship (PD/BD/129394/2017)

Summary

For practical reasons, surveys that aim for a large number of respondents tend to restrict themselves to closed-ended responses. Despite potentially bringing richer insights, open-ended questions pose significant challenges in extracting useful information while significantly increasing the analysis time. Nevertheless, automatic text analysis techniques could speed up the analysis of open-ended responses. Furthermore, open-ended questions in conjunction with closed-ended questions are likely to influence the closed-ended responses. Considering this, we pursued the following four objectives in this thesis, a. to analyse if the method of collecting qualitative data influences the survey responses, *b*. to develop an approach to extract open-ended responses from a survey and process the data, c. to compare the relative performance of the open-ended and closed-ended responses in analysing qualitative data, *d*. to develop a framework that measures attitudes while allowing respondents to choose their preferred type of question (closed- or open-ended).

This thesis analyses the suitability of using Topic Modelling to extract information from the open-ended responses to measure attitudes. As a case study throughout the whole thesis, we used questionnaires that collect information on the attitudes related to Autonomous Vehicles (AV). In this case study, alternative versions of the questionnaires that consider open- and/or closed-ended questions were presented randomly to respondents. Thus, two datasets were collected, a. 364 responses from India on the intention to use Shared AVs, b. 3002 responses from the USA on the intention to use AVs for commute trips.

Keywords

Topic Modelling, latent dirichlet allocation, supervised latent dirichlet allocation, likert scales, open-ended questions, travel behaviour research, model-based machine learning, bayesian estimation

Word Clouds for different open ended questions.