03/12/2024

Name Surname: Belkıs YÜCE

Student ID: 790824001

MIDTERM REPORT

This paper briefly explains the problem, method, data, and possible resources of the research. In this study, I aim to further the work I did in my master's thesis on the calculation of affective polarization. Therefore, first I will present the analyses I made within the scope of the thesis. Then I will explain the deficiencies of the analysis and how I will do the Python application.

Research Problem: Various calculation methods are suggested in the literature for calculating affective polarization, but I developed a new calculation method in my thesis with the idea that these cannot measure polarization comprehensively and accurately. There are two basic factors affecting affective polarization in the method I developed. One of these is the absolute value of the difference between distance and closeness scores to the identities. Another determining factor is the joint relative frequency, that is, in a crosstab the ratio of the number of people falling into each cell to the total number of people who indicated scores for both identities.

Research findings show that some groups have lower polarization than expected according to the literature. A possible reason may be as follows: Some of the identities (e.g. "Muslim") are well-known and socially acceptable so might be easily scored. However, some identities (e.g. "Deist") are not well-known and because of the fear of stigmatization, many participants may prefer to not score them. Therefore, high missing values in certain identity categories reduce joint frequencies and make affective polarization seem lower.

Method: Since I think my basic problem is high missing data in the joint frequencies of certain identity groups, I aim to use deep learning imputation to handle the missing data. Thus, using MLP, I will create a learning model only from the responses of people who respond to certain identity categories and reduce missing datas for those groups by producing new synthetic responses accordingly. According to the research findings I made during the preparation phase approximately 1900 sample sizes in total and 400 responses in the cross tables are sufficient for deep learning imputation. Keras, TensorFlow, or PyTorch seem to be libraries suitable for the structure of my data and my purpose in building the model.

Data: In this study, I will use the Turkish Faith and Religiosity Survey (TFRS) data, which I also used in my thesis. This survey was collected between December 2021 and May 2022, covering 35 provinces at the NUTS-1 level across Turkey. Accordingly, participants were asked to indicate how close or distant they felt to 21 social, religious, or political identity categories by scoring 0 (very far) to 5(very close).

Possible sources:

- MIDASpy GitHub Page via https://github.com/MIDASverse/MIDASpy
- TensorFlow Missing Data Imputation Tutorials via https://www.tensorflow.org/probability/api_docs/python/tfp/sts/impute_missing_values
- PyTorch Official Documentation via https://pytorch.org/tutorials/recipes/recipes_index.html
- Aurelien Geron. 2019. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems (2nd. ed.). O'Reilly Media, Inc.
- Murphy, Kevin P., "Probabilistic Machine Learning: An Introduction, 1st Edition" (2022). *eTextbooks for Students*. 498.