gesis Leibniz Institute
for the Social Sciences

hhu Heinrich Heine
Universität
Düsseldorf

Master Thesis

A “Certainty Score” for Question Attribute Prediction

GESIS maintains one of the largest archives of social science surveys, containing over 3000 surveys with over 50.000 questions. To allow easy searching and filtering within the question database, we want to enrich all questions with meaningful attributes. Researchers could, for example, be interested in filtering the topic of the question, what type of information it is asking for, or the answer type for the following question:

Question: Some vaccinations are compulsory. Do you consider this ..?
Answer Options: Perfectly acceptable
Fairly acceptable
Not very acceptable
Not at all acceptable
DK (Don’t know)
Not ascertained

Topic: Public Health
Asking for: an evaluation
Answer type: ordinal

First attempts to automatise the attribution prediction process have shown promising results; however, the precision of prediction models is not yet at a level acceptable for a productive system. We have a manually annotated subset of 6000 data points from the question database serving as the basis for our exploration.
As a trade-off between automation and quality, we want to explore the development of a "certainty checker" that separates very certain predictions from uncertain predictions [1] in the hope of automatising the annotation process at least partly.

We aim to:
- apply machine learning models to predict attribute values for the questions in the dataset
- develop and explore a "certainty mechanism" that identifies “safe” predictions for a productive system without manual check
- evaluate the certainty mechanism

The focus of the student's project can be varied according to the student's interests and background, but should be along the lines of the goals mentioned above.

Contact:
Andrea Papenmeier
GESIS - Leibniz Institute for Social Sciences
Unter Sachsenhausen 6-8, 50667 Cologne
E-Mail: andrea.papenmeier@gesis.org
Phone: +49 (221) 47694-229

Prof. Dr. Stefan Dietze
Heinrich Heine University Düsseldorf - Data & Knowledge Engineering
Universitätsstr. 1 Building: 25.12 Floor/Room: 01.39, 40225 Düsseldorf
E-Mail: stefan.dietze@hhu.de
Phone: +49 (211) 81-13785

Introductory Readings:
[1] Kannan, A., Givoni, I. E., Agrawal, R., & Fuxman, A. (2011, August). *Matching unstructured product offers to structured product specifications*. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 404-412).