Automatically Analyzing Open-Ended Survey Responses Using Statistical and Machine Learning Methods

Iyer, Devrath (School: Hamilton High School)

Surveys are an important part of behavioral research and design iteration. Current surveys are typically made up of closed ended questions(i.e. multiple choice questions), which have little information density and can introduce bias into the dataset. Open ended questions(i.e. free response questions) can provide much richer data, but are underutilized due to the amount of labor involved in grading them. The goal of this project was to create a system of computer progams that could automatically grade one type of open-ended response, the review(specifically Amazon reviews). The system first determines features of the product using the TF-IDF statistic. Keywords automatically generated by Amazon and TF-IDF were compared against a set of 10 manually selected goal words for 15 products, with TF-IDF performing better than Amazon's model 93% of the time. Second, any pronouns are resolved using neural network-based coreferencing and paragraph responses are broken into individual phrases by topic. Relations are extracted using dependency parsing and Open IE and are classified by topic. The system can then evaluate these relations and determine the specific word used to describe a feature and whether the respondent feels positively or negatively about any TF-IDF feature using sentiment analysis. VADER, an existing sentiment analysis algorithm, was enriched with dependency parsing to create a novel algorithm called MADER which improves its accuracy by 22%. Through these four steps the system can automatically generate quantitative insights into the way thousands of respondents feel about the features of a product in minutes given only the plain-text reviews.

Awards Won:
Arizona State University: Arizona State University Intel ISEF Scholarship