

Speaker:

Professor Matthias Schonlau

University of Waterloo

Title:

Semi-automated categorization of open-ended questions

Abstract:

Text data from open-ended questions in surveys are difficult to analyze and are frequently ignored. Yet open-ended questions are important because they do not constrain respondents' answer choices. Where open-ended questions are unavoidable, sometimes multiple human coders hand-code answers into one of several categories. At the same time, computer scientists have made impressive advances in text mining that may allow automation of such coding. Our preliminary work suggests automated algorithms do not achieve an overall accuracy high enough to entirely replace humans. We categorize open-ended questions using text mining algorithms for easy-to-categorize answers and humans for the remainder. Using an open-ended question in which respondents gave advice in a hypothetical situation at a doctor's office we find that up to half of the test data can be automatically categorized without sacrificing accuracy.