

Thesis Abstract

Saad Obaid ul Islam

Charles University, Saarland University

Title

Tackling Hallucinations in Chart Summarization

Abstract

Information visualizations like bar charts, line charts, and pie charts are a common way of communicating quantitative data. They are used to get important insights and make well informed decisions. Automatic Chart Summarization is the task to explain and summarize the key takeaways from the chart. Like other natural language generation (NLG) systems, chart summarization systems suffer from a phenomenon called hallucinations. Hallucinations occur when the system generates text that is not grounded in the input. In this research work, we try to tackle the problem of hallucinations in chart summarization. Our analysis shows that a lot of additional information is present in the training data that leads to hallucinations during inference. We also found out that reducing long distance dependencies and addition of chart related information like title and legends improve the overall performance of the system. Furthermore, we propose a natural language inference (NLI) based method to clean the training data and show that our method produces faithful summaries.