

ENHANCING TEXT-BASED CAUSAL INFERENCE OF
ONLINE REVIEWS USING MACHINE LEARNING
APPROACH

Sahana Sivaraj

A dissertation submitted in partial fulfilment of the requirement for Master of Science
degree in Bigdata Analytics

School of Computing

Informatics Institute of Technology, Sri Lanka in collaboration with
Robert Gordon University, UK

2022

Abstract

Causal Inference combines methodologies of evaluating the causal impact of treatment along with outcomes influenced by confounders where they need to be measured. This concept is an important research area topic in several areas such as computer science, economics, statistics etc. Recently, causal effects estimation on observations' data is growing as an upcoming research field since it has large size data sets in many domains especially measuring confounders from text become an interesting research topic. For example, how a product review majorly impacts its sales increment? As the review content is majorly in a text format this will be a bit much challenging to measure causal effects of text since it has a high dimensional structure. Various causal impact estimating methods for text have arisen in the wake of the fast evolution of machine learning. But still, it faces the selective biased problems where it produces invalid causal effects estimations. As mentioned in the example the major reviews systems analysis the impact of sales increments based on numerical ratings where it ignores important facts hidden from review content.

To overcome above mentioned issues, by combining causality with machine learning are sufficient for forecasting and understanding the cause and effect of the results. This research is utilized systematic approach to identify which causal machine learning methods are used to enhance the causal inference problems related to textual reviews system to validating how it is impact on the products sales/demand using amazon platform as an example. This research also focuses on the combine use of advanced causal models like Double Machine Learning with machine learning approaches to conduct causality analysis using amazon online review data.

Key Words – Causal Inference, Machine Learning, Double Machine Learning, Online Reviews System, Natural Language Processing