

Thesis Project Portfolio

Analysis of Shortcut Learning Features in Natural Language Processing

(Technical Report)

**Machine Translation Technology: The Advantages and Limitations of Machine Translators
in the Academic Community**

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Wan Li

Spring, 2022

Department of Computer Science

Table of Contents

Sociotechnical Synthesis

Analysis of Shortcut Learning Features in Natural Language Processing

Machine Translation Technology: The Advantages and Limitations of Machine Translators in the Academic Community

Prospectus

Sociotechnical Synthesis

Introduction

The two projects identified in this research proposal consist of the technical paper, which focuses on shortcut identification in machine learning models, and the STS research, which explores the effect of machine translation (MT) in education. These are connected through machine learning - both projects have a central focus around machine learning and its application in society. The technical portion aims to improve these machine learning models by proposing a systematic approach to identify shortcuts, which would further improve the accuracy of the models. The STS portion investigates the effect of MT in society, where MT is a specific application of machine learning. Together, experience is gained regarding improving current technology and exploring how that technology would ultimately affect society.

Technical Research Paper Summary

Machine learning has advanced greatly in the past few years, but there are still many limitations. Many problems related to difficult machine learning problems are symptoms of shortcut learning. Shortcut learning, as Geirhos et al. puts it, are “decision rules that perform well on standard benchmarks but fail to transfer to more challenging testing conditions, such as real-world scenarios” (2020). Essentially, shortcuts reveal a mismatch between the model’s intended solution and the learned solution. Interpretations shed light onto what logic a model uses to solve a problem. The technical deliverable ultimately aims to systematically identify possible shortcuts a model may take based on the training dataset. This goal includes observing dataset bias and combining it with interpretation results to identify the possible shortcuts, as well as how they relate to the dataset as a whole. “Backdoor shortcuts,” or shortcuts that cannot be

identified easily by humans, are the focus of the systematic identification. Mitigation of the shortcuts will be considered, but is not in the scope of this project. Identifying shortcuts in a model will allow researchers to develop better datasets and more accurate machine learning models that avoid superficial correlations.

STS Research Paper Summary

Machine Translation (MT) is the translation of text from one language to another using software. It is a relatively new field, with research starting in the 1950s and continues today. As MT becomes more refined, it has noticeable effects on the global academic community and society as a whole. The STS research paper aims to explore MT's limitations and ultimately understand the effects of MT on academics. With all the advantages and limitations of MT in society, what impact does Machine Translation technology have on education? The evolution of MT and how it drives the academic community will be discussed in the context of technological determinism as well as technological momentum, which observes MT in a timeframe. The research will uncover the connections between MT and academic integrity, MT's relation to accessible education for all, and how MT may change language requirements for higher institutions. Studying the effects of MT on society is important because it is continuously improving and changing the world of academia. If its implications are not addressed through policy and awareness, societies will not know how to move forward and adapt to this quickly evolving technology. Understanding MT's history and how it affects societies through technological determinism will allow us to create informed decisions in the present and future regarding MT.

Conclusion

Working on both the technical project and the STS research paper simultaneously has increased my understanding and appreciation of how technology affects society. As engineers, we often think solely of how to solve the problem, but we hardly stop to think about how these solutions may change society. If I had worked on just the technical project, I would not have considered the effect improving machine learning models would have on varying society, whether advanced or developing. If I had only worked on the STS research paper, I would not realize why societal effects connect to my work as an engineer. Working on them together has given me appreciation for how STS plays a large role in the everyday lives of engineers. We must stay diligent in understanding the results of our work and how that may impact others. Especially with machine learning advancing so quickly over the years, we must improve our technology yet make sure it does not cross a line - we want technology to aid us, not rule us. Where that line is can only be determined by those that understand both technology and their effects on society.