Technology Review

Nathaniel Rupsis (NetId: nrupsis2), CS410 - Text Information Systems

Topical overview of text information models, methods, and techniques for Education

Introduction

Over the course of past several decades, we've seen an exponential growth of data creation, dissemination, and availability. According to Statista.com [1] the volume of data created world wide has increased 30 fold since 2010, and is forecasted to triple over the next 4 years. Because of this, we have more information at our fingertips then ever before. However, too much of a good thing can lead to some unintended consequences. This paper will briefly summarize some areas of research which aims to help the educational efforts to provides highly relevant, useful, and summarized information for learning.

Utilization of Big Data Methods

In this overview, we'll be examining two distinct, but similar research projects that focus on providing complementary information to web based learning.

The first research project is AXIS (Adaptive eXplanation Improvement System)[2]. The aim the of the AXIS system is to provided high quality explanations of educational material via a Human Centric feedback model. This feedback cycle (aided with some machine learning), continuously improves the quality of resources via user feedback prompts. AXIS is composed of 2 main components, a data collection interface (learnersourcing) and an explanation selection policy (a "multi-armed bandit"problem model which selects which resource to deliver). This 2 part yin and yang system delivers problem solutions (or other relevant learning material) to the user, then immediately prompts for feedback.

This feedback loop is what drives the recommendation engine. The AXIS system uses Thompson sampling (a Bayesian algorithm). This Thompson model stores a probability for each item, and makes it's selection based on the highest perceived reward (I.e, a high rating from the user). Once the selection is made, the parameters are used as a prior. The Likelihood used in the model is Bernoulli distribution, and is updated based on each reward observation (feedback submitted). This paper unfortunately doesn't provided any formal equations explaining the underlying models used. As far as I'm aware, the project is still active, and if you want to participate, you can signup here: http://tiny.cc/useaxis.

The second research paper is in a similar vain as AXIS, where it programmatically adds meta information to help aid in learning. Project Tutorons [3] from UC Berkeley aims to help provide on demand, context relevant, information to students as they learn. The two mediums used in the research paper are CSS selectors, and some basic Unix commands.

The basic premise for Tutorons, is dynamic parallel requests made to backend language supporting servers, each called a "Tutoron". The information that is send back to the user is called a "Micro-explanation" which is either a natural language, or code specific explanation of what the requesting object does.

The main "big data" method used in Tutoron, is the topic analysis parsing. While the parsing, the first phase is to extract a block of code, and try to create some sub-context elements (for example, html elements). Once the candidates have been identified, it's passed to language specific parser (CSS vs Javascript, etc), from there it's parsed with a syntax to check if it fits a specific grammar (The paper doesn't mention what happen if it can, or can't find a match). From there, it dispatches a request to grab information about the parsed element.

The results of this research project were mixed, `wget` achieved 95% precision, and 65% recall. Whereas `regular expressions` achieved 70% precision and 14% recall. The authors noted that the failure for regex is the context specific nature of implementations. Finally the paper noted that for future work, they are interested in rule-based and learned weight models to help increase the overall accuracy.

Conclusion

References

[1] Holst, Arne. "Data Created Worldwide 2010-2024." Statista, 7 July 2020, www.statista.com/statistics/871513/worldwide-data-created/.

[2] Williams, Joseph Jay, et al. "Axis." Proceedings of the Third (2016) ACM Conference on Learning @ Scale - L@S '16, 2016, doi:10.1145/2876034.2876042.

[3] Head, Andrew, et al. "Tutorons: Generating Context-Relevant, on-Demand Explanations and Demonstrations of Online Code." 2015 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), 2015, doi:10.1109/vlhcc.2015.7356972.