Inferring Properties of Causal Structure from Student Essays

Simon Hughes[1]*, Peter Hastings[1], Anne Britt[2],
Dylan Blaum[2], and Patty Wallace[2]

[1] DePaul University, Chicago, Illinois
[2] Northern Illinois University, DeKalb, Illinois

Abstract. In the US in particular, there is an increasing emphasis on the importance of science in education. To better understand a scientific topic, students need to compile information from multiple sources and determine the principal causal factors involved. We describe an approach for automatically inferring the quality and completeness of causal reasoning in essays on two separate scientific topics. We present a novel machine learning method using a two-phase approach for detecting causality. For each core essay claim, we initially trained a word window based tagging model to predict which individual words belonged to that claim. Using the predictions from this first set of models, we then trained a second stacked model on all the predicted word tags present in a sentence to predict inferences between essay claims. The results indicate we could use such a system to provide explicit feedback to students to improve reasoning and essay writing skills.

Keywords: Reading, Argumentation, Causal Inference, Machine Learning, Natural Language Processing, Essay Grading

1 Introduction

test: I am going to cite [1] et al. **TODO**

2 Related Work

3 Manual Essay Annotation

Two document sets containing 5 unstapled, single-sided sheets were developed to assess students ability to integrate information to develop an understanding about a scientific phenomen: coral bleaching or skin cancer. The documents were

* The assessment project described in this article is funded, in part, by the Institute for Education Sciences, U.S. Department of Education (Grant R305G050091 and Grant R305F100007). The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

prepared from reputable sources (the NASA earth observatory, the US Geological Survey, and online science textbooks), and each started with some short background material to provide framing, necessary vocabulary, and background knowledge required to understand the rest of the material. To ensure that the students could more easily pull different pieces of information from different sources, the documents were not stapled together and could be read in any order. The sources were also compiled such that the students needed to combine information from multiple different source materials to fully answer the question.

Prior to development of the document sets, a causal model of each scientific phenomenon was created (see figure 1) that visually describes the causality detailed in the source documents relevant to answering the essay question. This causal model is a coherent series of claims connecting initiating factors (e.g., increased trade winds, warmer waters in east, increased salinity) to the to-be-explained outcome (TBEO) (e.g., coral bleaching).

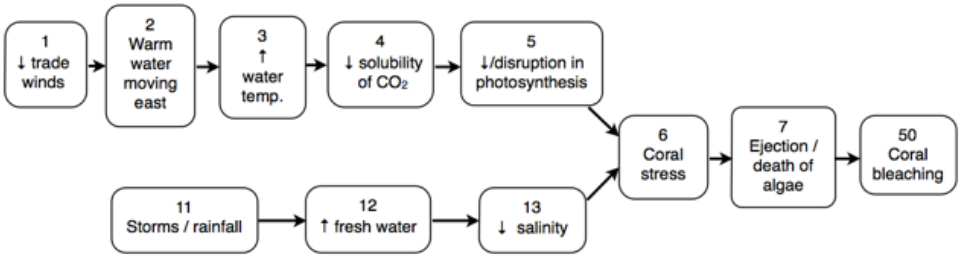

Fig. 1. The Coral Bleaching Causal Model

Each student was provided with the essay prompt, and asked to answer the question using the source material provided. 105 middle and high-school students were assigned the 2 essay questions. Following completion, the essays were then annotated by two different annotators according to how well they aligned with the corresponding causal model. Each word was tagged with its corresponding concept from the model, if applicable, and links were created between tagged sequences representing the causal relationships. The inter rater reliability was high ($\kappa = 0.85$), and the technique proved useful in determining the level of coherence, and the quality and completeness of the causal reasoning within each essay.

4 The Tagging Problem: Identifying Concept Codes

5 Detecting Causality

6 Evaluation

The ideal essay was one which identified all the causal concepts, and illustrated the correct causal chains and in the correct order. To this end, multiple metrics were defined to evaluate the relative completeness of each student's answer compared to the causal model for the topic. TODO

7 Conclusions

References

1. Vladimir N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York, Inc., New York, NY, USA, 1995.