Modeling Biological Processes for Reading Comprehension

Jonathan Berant, Vivek Srikumar, Pei-Chun Chen, Brad Huang Christopher D. Manning[1], Abby Vander Linden, Brittany Harding[2], and Peter Clark[3]

[1]**Stanford University, Stanford**
[2]**University of Washington, Seattle**
[3]**Allen Institute for Artificial Intelligence, Seattle**

Review by: Jake Pitkin

INTRODUCTION

Machine reading is a subfield of natural language processing with the goal of developing programs that can process text, learn about the world, and make educated decisions. Most work in machine reading uses an approach called macro-reading. Macro-reading uses web-scale corpora to extract facts that appear a redundant number of times. In this work the authors explore micro-reading: given a single document with little to no redundancy can the system answer a comprehension question?

The authors focus on the task of taking a paragraph describing a biological process and to answer questions that require deep understanding. They hypothesize that in order to accomplish this, rich structures that represent the process must be predicted. That is, we must identify and model the interactions betweens the events and arguments of the process. They demonstrate that using these predicted structures provides a substantial improvement over shallower representations of the corpus and the questions.

STRENGTHS OF THIS APPROACH

When given the task of answering a factoid question, macro-reading has been an approach shown to get decent results (Etzioni et al., 2006; Carlson et al., 2010; Fader et al., 2011). Given a large corpus, a knowledge base can be extracted to answer factoid questions where the answer exists a redundant number of times in the corpus.

However these approaches don't apply to the micro-reading setting where we are reading a single document with the goal of answering a question than requires more complex reasoning than answering factoid-style questions. This work is strong in that is identifies that to answer dependency, temporal, and true-false questions, the system needs a deeper knowledge about the interactions between the events and arguments of the process. It is the first of it's kind to construct a system that predicts rich structured representations of this knowledge and to use it to answer difficult reading comprehension questions. The authors demonstrate that structured representations have sizable accuracy gains over shallower representations for this task. Given the golden structure, their system gets 76.7% of questions correct over the best shallow representation, SYNTPROX which achieves 60%.

Good work was put into constructing quality questions and answers as well as annotated data for training and testing the system. 200 paragraphs were hand-picked describing biological processes. From these paragraphs, 585 questions were created by a biologist ensuring the questions focus on reasoning about the relations between those events and entities. Two

annotators who were also biologists evaluated 326 questions at random and agreed 98.1% of the time. An important note is questions were authored independently of structure annotations. The combination of all these elements makes for a fair empirical evaluation with no bias towards their system.

EXPERIMENTS

I found the experiments and error analysis to be strong in this work. The simple choice for a baseline would be random choice which would be 50% given there are two options. Instead the authors explore results using lexical overlap, text proximity, and syntactic proximity as baselines. This is a strong way to sell the point that rich structures provide substantial improvements over shallow representations. Given the golden structure, they find a 16.7% accuracy improvement over the best baseline.

In addition to determining structured representations are vital to understanding difficult questions, they demonstrate the importance of the structure prediction accuracy. Gold structures see an impressive 10% gain in overall accuracy. Hinting at the proposed method in this work is sound and structure prediction improvements should be the focus of future work.

FURTHER DEVELOPMENT

The authors show that their system sees substantial accuracy gains as the structure prediction accuracy increases (as one would naturally expect) and more data would improve performance. Annotation is not cheap, especially for this task which requires a background in biology. There exists a lot of biology literature that could be used if low supervision techniques were developed. Follow-up work could try using a bootstrapping method to leverage an unannotated corpus in a semi-supervised setting. Seeds could be obtained from the PROCESSBANK dataset from this work.

The task introduced in this work provides two possible answers to the question and the system decides between the two. I found this to be a fair task to demonstrate the new approach presented in this work. But it would of course be ideal if the system could answer questions without providing it possible answers. Perhaps a system could be developed that predicts a collection of candidate answers to the question. Then, these candidates could be evaluated by the presented method to predict a final answer.

CLOSING THOUGHTS

I found this to be strong work with a well-defined objective. The authors present a good case for the necessity of expressive structures in micro-reading tasks that attempt to get at the deeper meaning of short corpora. More formally, if we hope to have efficient systems that answer questions based on a small corpus, we must directly model the interactions between events and arguments of a process. As the authors state, event and argument relationships appear in a wide range of domains and this work should generalize well.

Overall I found this work to be significant as it produced a novel method to a challenging problem. What is equally as important is it also provided a new task to evaluate micro-reading methods that answer comprehension questions along with a new dataset PROCESSBANK. This provides a strong base for future research into micro-reading, which is an essential compliment to macro-reading in the field of machine reading.

REFERENCES

Oren Etzioni, Michele Banko, and Michael J. Cafarella. 2006. Machine reading. In *Proceedings of AAAI*.

Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R. Hruschka Jr., and Tom M. Mitchell. 2010. Toward an architecture for never-ending language learning. It *Proceedings of AAAI*.

Anthony Fader, Stephen Soderland, and Oren Etzioni. 2011. Identifying relations for open information extraction. In *Proceedings of EMNLP*.