COMP SCI 397
Course Project Final Report
Tracking State Changes in Procedural Text

Danilo Neves Ribeiro
dnr2876
daniloribeiro2021@
u.northwestern.edu

William Hancock
todo123
williamhancock2022@
u.northwestern.edu

Abstract

In this project we worked on the recent AI2 question-answering benchmark called ProPara. This data set is comprised of procedural text paragraphs covering different topics (e.g. photosynthesis) and the objective is to keep track of how state of entities involve through time (e.g. water or light gets absorbed by the plant). Our approach follows the ProGlobal model from the original ProPara paper. We show how our results compare to the published results and also compoile an error analysis on the results of the original paper.

1 Introduction

Answering questions about paragraphs that describe processes is still a challenging task for machine reading comprehension systems. This genre of text is pervasive (e.g. manuals, recipes, road safety rules, scientific protocols, etc.) and understanding them often requires keeping track of how the worlds state evolve over time. For instance, consider the paragraph describing photosynthesis in Figure 2. If the system is asked the question: "Where is sugar produced?", it is expected to answer "In the leaf". To answer the question, the system needs to infer the state changes of each entity in the paragraph and the causality between such change events (which are often implicit, making this a challenging task). The dataset is further detailed in the following section.

2 Data Set

To evaluate our system, we use the ProPara procedural text benchmark which contains 488 crowdsourced paragraphs and 3100 sentences total. This data set is comprised of procedural text paragraphs covering different topics together with a human annotated table that describes the state (location and existence) of entities in this paragraph. Figure 2 shows an instance of the training data which constitutes of a paragraph about photosynthesis and the annotated state change grid. The state change grid contains information about where an entity (e.g. water or light) is at each step. Note that "?" indicates the location is unknown, and "-" indicates the entity doesn't exist during that step.

		Participants:				
Paragraph (seq. of steps):		water	light	CO2	mixture	sugar
	state0	soil	sun	?	-	-
Roots absorb water from soil						
	state1	roots	sun	?	-	-
The water flows to the leaf.						
	state2	leaf	sun	?	-	-
Light from the sun and CO2 enter the leaf.						
	state3	leaf	leaf	leaf	-	-
The light, water, and CO2 combine into a mixture.						
	state4	-	-	-	leaf	-
Mixture forms sugar.						
	state5	-	-	-	-	leaf

Time

Figure 1: ProPara participant state change grid.

3 Model and Implementation

In this project we implement a model similar to ProGlobal, which was introduced in Dalvi et. al. The key feature of this model is that it predicts the state of *all* participants at each timestep (even if the participant is not mentioned in a sentence). Therefore the persistence of the states are tracked by the model itself.

4 Evaluation and Results

During testing the system will 4 categories of questions from the output state change grid:

1. What are the Inputs? That is, which participants existed before the procedure began

2. What are the Outputs? That is, which participants existed after the procedure ended?

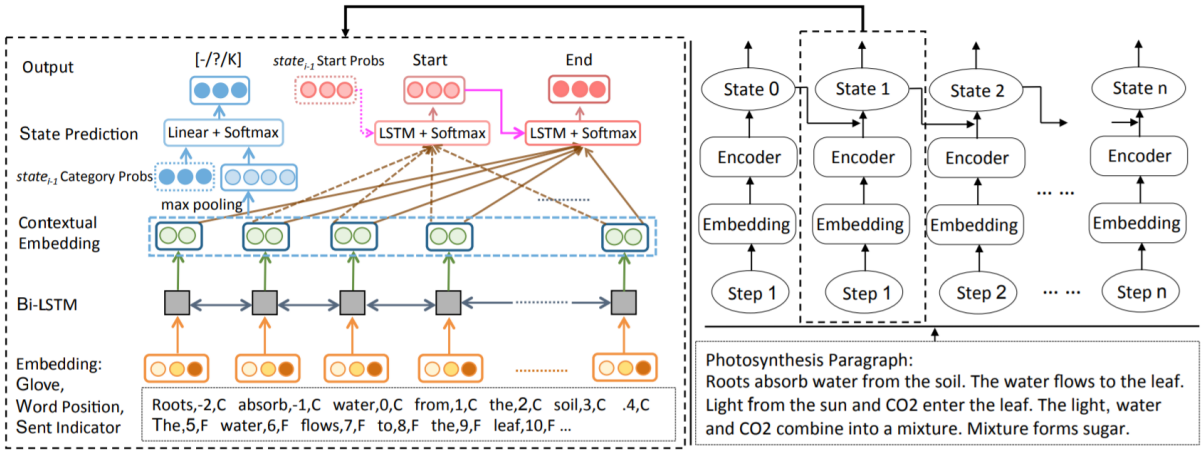

Figure 2: Diagram describing the ProGlobal model. The input of the model is the whole paragraph, plus features indicading the current sentence and the distance to the participant. Previous step predictions are fed to next step prediction

Model	Precision	Recal	F1
ProLocal	**77.4**	22.9	35.3
ProGlobal	46.7	**52.4**	49.4
ProStruct	74.2	42.1	**53.7**
Our Results	76.9	25.0	37.7

Table 1: Results on test set.

3. What are the Conversions? That is, which participants were converted to which other participants?

4. What are the Moves? That is, which participants moved from one location to another?

Note that these questions are templated, meaning they can be deterministically answered using the output state grid. The evaluation code was made available at https://github.com/allenai/aristo-leaderboard/tree/master/propara by the AI2 team. The code outputs precision, recall and F1 score for each question category.

5 Conclusion

TODO