Supplementary Material: Modeling Biological Processes for Reading Comprehension

Jonathan Berant,[*] **Vivek Srikumar,**[*] **Pei-Chun Chen, Brad Huang and Christopher D. Manning**
Stanford University, Stanford

Abby Vander Linden and Brittany Harding
University of Washington, Seattle

Peter Clark
Allen Institute for Artificial Intelligence, Seattle

1 Features for Trigger Classifier

For trigger classification, the following binary features are extracted given a trigger candidate:

1. The WordNet synsets of the candidate
2. The Nomlex nominalization type
3. Whether the trigger is an auxiliary verb
4. Clustering of trigger candidate using WordNet to one of the direct hyponyms of the synset *Entity#n#1*
5. Levin verb classes of trigger candidate
6. Adverbial modifiers of trigger candidate
7. Lemma and POS tags of preceding two words
8. Lemma and POS tags of following two words
9. Dependency path to root of sentence
10. POS tag of dependency parent conjoined with the POS of trigger candidate and the connecting dependency edge relation
11. For each dependency child, POS tag of dependency child conjoined with the POS of trigger candidate and the connecting dependency edge relation
12. Indicator for the dependency parent being a nominalization
13. Indicator for any dependency child being a nominalization
14. Indicator for trigger candidate being a nominalization whose dependency parent is a verb
15. Indicator for the trigger candidate being in a gazetteer of biological processes compiled from Wikipedia

2 Features for Argument Identifier

We extracted the following features for our argument identifier, given an argument candidate and the corresponding trigger:

1. The syntactic category of the argument candidate
2. The POS tag of the trigger

3. Conjunction of trigger POS tag and argument category
4. Indicator for the the argument node containing a sentence category
5. Indicator for the existence of a dependency relation between the trigger and argument
6. Dependency path from trigger to argument
7. Length of dependency path from trigger to argument

3 Features for Joint Model

Event-Argument features For a given trigger and argument candidate, we extract the following features:

1. The head word, its lemma and POS tag,
2. The parse tree node that covers the argument,
3. The subcategorization frame,
4. Indicator for whether the argument contains an SBAR node,
5. Dependency path from the trigger to the head of the argument,
6. Length of the dependency path,
7. Path from argument node to the root of the constituency tree
8. Indicator for whether the argument is before or after the trigger,
9. Number of tokens between the trigger and the argument,
10. Lemma of head of argument
11. Lemma of head of argument conjoined with POS tag of trigger
12. Lemma of head of argument conjoined with lemma of trigger
13. Lemmas of words between the trigger and argument candidate
14. From the verb (PropBank) and nominal (NomBank) semantic role annotation, an indicator for whether the trigger is an predicate and the argument candidate is an argument of that predicate,
15. For the previous case, the label of the argu-

---

[*] Both authors equally contributed to the paper.

ment, and

16. The previous two features, when the argument heads match.

Event-Event Relation features Given a pair of triggers, we extract the following features:

1. Lemmas of two words preceding and following each trigger
2. Lemmas of both triggers
3. Indicator for whether the two triggers have the same lemma
4. Word and sentence distance between the triggers (word distance is binned into buckets of <5,6-7, 8-10, 11-15,16-30, >31)
5. The determiner of the trigger, if any
6. Conjunctions of the following features:
   - Lemmas and POS tags of the triggers and
   - Cluster identifiers if both triggers are contained in a cluster, using EX-CHANGE clustering.
7. Adverbial modifiers for triggers
8. Lowest common ancestor of triggers in constituency tree, if it exists
9. Dependency path between the triggers
10. Length of dependency path between the triggers
11. Indicator for whether first trigger dominates the second in the dependency tree
12. Indicator for whether second trigger dominates the first in the dependency tree
13. The child of a *mark* dependency relation, if one exists
14. Preposition lexeme, if in a prepositional phrase
15. Indicator for whether triggers share a dependency child
16. For each trigger, indicator for whether the trigger is an SRL predicate
17. Indicator for whether triggers share an SRL argument
18. Indicator for whether the triggers are adjacent in the paragraph
19. Indicator for whether triggers are adjacent and which trigger is first
20. Words between triggers
21. Indicator for whether first trigger is a noun and the first word in the paragraph