System Description for SAIC Entry at RTE-6

Demetrios G. Glinos
Science Applications International Corporation
demetrios.g.glinos@saic.com

Abstract

SAIC participated in the RTE Main Task (Summarization scenario) at TAC 2010. The SAIC entry implemented a rule-based analysis approach involving deep parsing, WordNet, and a propositional pattern matching strategy for entailment determination. Only a first working version of the implementation was available for use for the challenge, with corresponding limited results.

Introduction

SAIC participated in the Recognizing Textual Entailment challenge main task at TAC 2010 [1]. This task simulated a summarization scenario in which, given a corpus of articles, a set of hypotheses, and a set of candidate sentences retrieved by the NIST search engines, systems were required to determine which of the hypotheses were entailed by their associated candidate sentences. The corpus was divided into 10 topics, each with 10 articles from which the candidate sentences were selected. There were 243 hypotheses in the test set, each of which was paired with multiple candidates. Overall, there were 19,972 hypothesis-candidate pairs, of which 945, or approximately 4.7%, involved entailment. For purposes of the challenge, a candidate is considered to entail the hypothesis if a reasonable person would believe that the hypothesis is more likely true given the candidate than without it. In other words, the candidate must actively contribute to the truth of the hypothesis.

System Description

Figure 1 shows the processing flow for the SAIC entry that participated in RTE-6. The transformations shown are performed for each hypothesis-candidate pair *de novo,* without regard to the results for any other pair. Syntactic parsing of hypothesis and candidate sentences, and of the article in which the candidate appears, is performed using the Stanford PCFG parser [2]. Parser output is then examined to extract a set of atomic propositions, each containing an un-nested predicate and set of arguments and modifiers for each original sentence. Complex original sentences are simplified by the extraction of propositions for relative and subordinate clauses, appositions, compound clauses, verb coordinations, and direct and indirect quotes. Thus, even hypothesis and candidate sentences can each be decomposed into multiple propositions. The proposition phrases are then assigned semantic roles using the PropBank [3] annotation guidelines. The named entity recognition component extracts the named entities from each role-assigned chunk and generates name variations based on titles for persons, organizational suffixes and acronyms, and noun-noun constructions, for use in pattern matching.

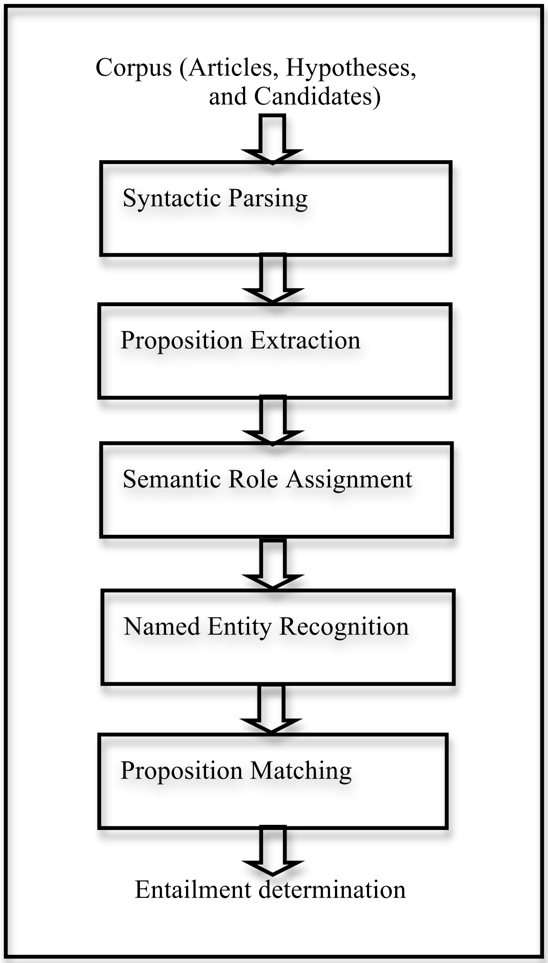

Figure 1. RTE Processing Flow

Entailment is determined using a proposition matching algorithm that searches the database of propositions for a proposition that matches the form of the hypothesis, first from among those for the candidate sentence, then from the article, and finally from among all the articles for the topic. A proposition is considered to match if it contains a synonymous predicate and contains a similar set of arguments. We use the WordNet [4] synonym and hypernym relations to determine the similarity of verbs. For arguments, we compare their embedded entities using the previously generated name variations and aliases. Where arguments do not correspond, the algorithm searches for "is" and "is-a" propositions in the database that would establish the required correspondence. The algorithm is constrained in its operation always to include one of the propositions from the candidate sentence as part of the set being considered for a match. This constraint enforces the requirement that the candidate contribute to the veracity of the hypothesis.

Performance

The system described did not perform well against the test set. Overall micro-averaged precision was 7.92, recall 21.69, and f-measure was 11.60. These results can be ascribed to several factors. First, development resources were limited, so that it was not possible to fully develop the proposition extraction component to include a more robust set of derived propositions, such as for genitive modifiers and the generation of predicates using antonyms. The inclusion of a coreference resolution component for nominal and pronominal references would also no doubt have improved performance.

The approach taken was enticing for the reason that the pattern matching algorithm produces a straightforward derivation for each entailment determination, providing a listing of the propositions used in establishing the matching. In examining the operation of the system in detail, however, it appears to us that the primary cause of poor precision was error in the decomposition of complex sentences into their atomic propositions. Post hoc analysis shows that while many valid propositions are extracted, the algorithms also generated a number of propositions that do not make sense and which, for technical reasons, match the hypothesis patterns. Future work is indicated both in completing the components indicated above, and also in pruning the proposition set upon which the pattern matching algorithms operated.

A typical run of the SAIC system processed all 19,972 hypothesis-candidate pairs, including all necessary parsing, in 7,450 seconds, corresponding to an average of 0.373 seconds per pair, running on a hardware platform consisting of dual core 2.66 GHz processors with 4 GB RAM memory.

References

[1] NIST 2010. Details concerning TAC may be found at http://www.nist.gov/tac/about/index.html.

[2] Dan Klein and Christopher D. Manning. 2003. Accurate Unlexicalized Parsing. *Proceedings of the 41st Meeting of the Association for Computational Linguistics*, pp. 423-430.

[3] Palmer M, Kingsbury P, Gildea D (2005). "The Proposition Bank: An Annotated Corpus of Semantic Roles". *Computational Linguistics* **31** (1): 71–106.

[4] Fellbaum, C. ed. 1998. *WordNet: An Electronic Lexical Database*. Cambridge, Mass.: The MIT Press.