On the Relationship of AURA Question-Answering and Textual Entailment

Working Note 31

Peter Clark, peter.e.clark@boeing.com, July 2009

Introduction

There are some important relationships between question-answering in the AURA question-answering system and recent NLP work in Recognizing Textual Entailment (RTE). This document aims to articulate these relationships, as they provide some insight for AURA's future question-answering pipeline.

To make the comparison easier, biology-related RTE examples are used, and we assume a biology context throughout.

Recognizing Textual Entailment (RTE)

The RTE task is, given some text T, decide whether a hypothesis H follows from T. For example:

T: The semifluid substance within the membrane of a prokaryotic cell is the cytosol.
H: A prokaryotic cell has a semi-fluid region consisting of cystol.
ANSWER: TRUE (H is entailed by T)

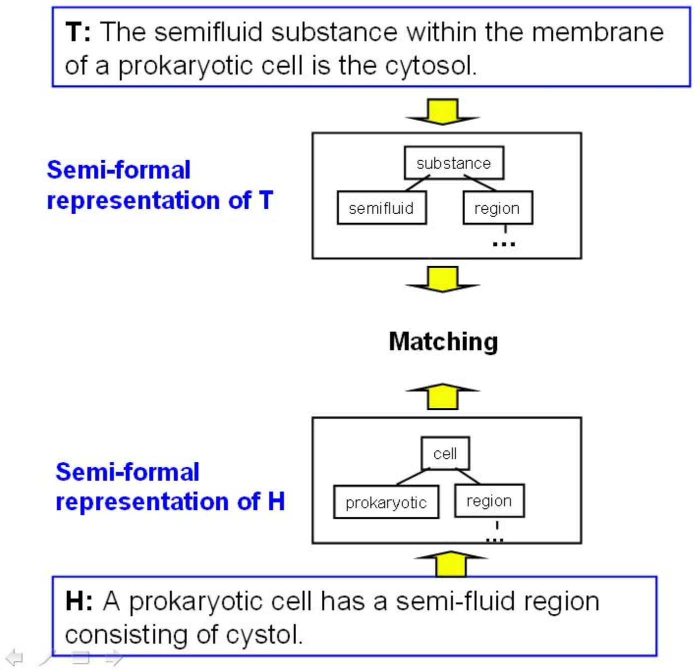
T: Peroxisomes contain enzymes that transfer hydrogen from various substrates to oxygen, producing hydrogen peroxide as a by-product.
H: Peroxisomes can catalyze reactions involving hydrogen peroxide.
ANSWER: TRUE (H is entailed by T)

In RTE, the notion of "follows" or "entailed" is not precisely defined. It is taken to mean: if a person (with common understanding of language and common background knowledge) would reasonably conclude H, given T, then H is said to be entailed by T. Thus textual entailment is defined according to human judgement, and thus includes a degree of subjectivity.

RTE systems typically work as follows:

1. Convert the T sentence to a semi-formal representation
2. Convert the H sentence to a semi-formal representation
3. Match the two semi-formal representations (see of H is "more general than" or "subsumes" or "is contained in" an elaboration or transformation of T)

For example, sketching the transformation sequences mentioned above:



The semi-formal representations produced in steps 1 and 2 might simply be two dependency (parse) trees (e.g., (MacCarney et al., 2006)), a semi-formal logical representation (e.g., (Clark et al. 2008)), or a full logical representation. Similarly the matching process in step 3 can vary in formality, from matching dependency trees (e.g., (MacCarney et al., 2006)) to full logical deduction (e.g., (Clark et al. 2008)). Typically the matching process involves use of background knowledge, in particular WordNet and the DIRT paraphrase database.

Note that "matching" is a somewhat imprecise term for the process of comparing the representations. In particular, the word "matching" misleadingly implies a symmetric operation (if X matches Y then Y matches X), whereas in reality the process is not symmetric: T must be (in some sense) "more general than"/"implied by"/"cover" H, and not vice versa, e.g., "A big cat is black" implies "A cat is black", but not vice versa[1].

Although it might seem RTE is restricted to answering yes/no questions, in fact it can easily be used for "find the value" questions by including a variable-like object in H. For example, to answer a question "When was Mozart born?", the H sentence "Mozart was born on a date" would be used, and (if H was entailed) the element/string in T that matched "date" would be the answer.

In the above examples, T is a single sentence. However, in general T may be a whole paragraph or (in principle) even a whole text book. For example, this year's RTE

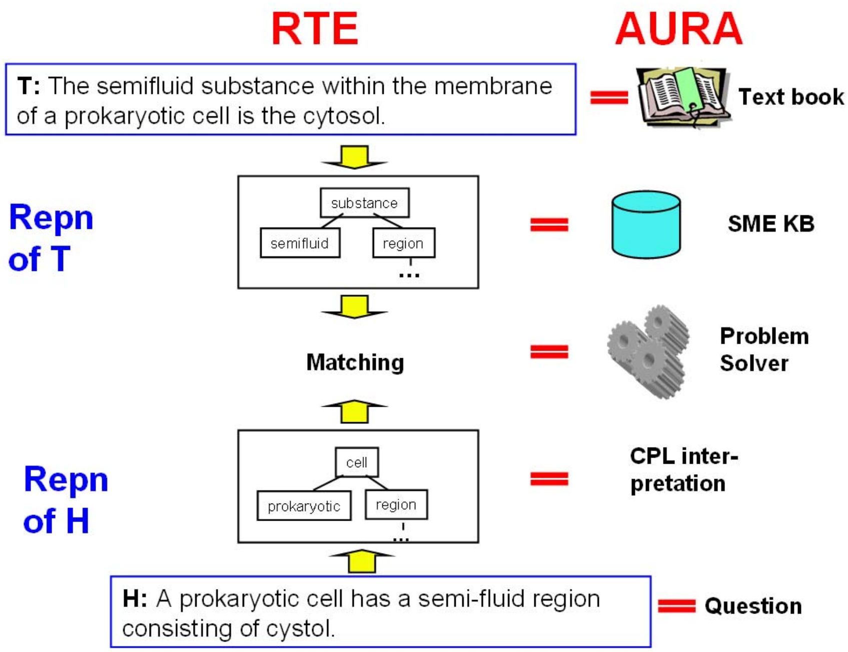
---

[1] In fact it is more complicated still, as negation, quantification, and modals can all reverse the direction (polarity) of the entailment matching (Nairn, Condoravdi, and Karttunen, 2006).

competition involves a new "RTE search" challenge where T is a set of 10 newspaper articles (and H is a single sentence).

RTE and AURA's QA Process

The two RTE examples above are interesting as they are drawn directly from the AURA final evaluation: The T sentences are from text books, and the H sentences are final evaluation "Is it true...?" exam questions (#2a and #5a). Thus the mapping between RTE and AURA is as follows:



In some ways AURA has an easier task than RTE: In RTE, automatically generating the initial representations of T and H is itself error-prone, even before matching commences. In contrast, the representations of T (i.e., the KB) and H (i.e., the CPL interpretation) in AURA are already available and can be considered likely to have high fidelity.

Lessons from RTE

There are some interesting characteristics of RTE that may have relevance to the future design of AURA.

First, RTE systems typically use fairly underspecified representations of T and H, e.g., not committing to word senses until matching, and often just use a simple parse of the sentences as the representation. The corresponding lesson for AURA may be to not over-engineer the initial question (CPL) interpretation. In RTE, fairly uncommital representations are used, to be subsequently aligned by the matcher. A similar approach may be appropriate in AURA. It might be CPL needs to produce a less, rather than more, fully disambiguated interpretation.

Second, in RTE the T sentences rarely deductively imply the H sentence. The two examples shown earlier are good illustrations of this. The cause of this is the variability in how language can be interpreted. Although this problem is removed for interpreting T (as the SME hand-builds the correct interpretation of the text book), it is not removed for interpreting the H (the question). This suggests that in AURA, at least in biology, a purely deductive problem-solving approach is unlikely to be sufficient, at least when the CPL questions (H sentences) are close to the original English.

Third, the bulk of the work in RTE occurs in the matching. In particular, it is the matching step, rather than T and H interpretation, where paraphrases and linguistic knowledge are heavily exploited to align the texts and coerce the question (H) into a form matching the KB. This suggests that, architecturally speaking, the use of paraphrases and linguistic knowledge may better be considered part of problem-solving rather than initial question interpretation, as we seek to introduce those resources into AURA.

This note does not intend to imply that RTE technology is “the answer” to performing better question-answering in AURA. In fact, the performance of RTE systems is still fairly poor: for example, the average system score in the 2008 RTE4 competition was 57% (two-way test) compared with 50% for random guessing (Giampiccolo et al, 2008). It thus seems unlikely that a pure RTE approach to biology question-answering would do well (i.e., see which questions were entailed by the text-book directly), although it might be an interesting experiment. Rather, AURA has some significant advantages by having the SME transduce complex text-book language into coherent knowledge structures, and having a reasoning engine that can combine disparate information together (something that RTE systems have difficulty with). Thus a more fruitful approach might be to integrate entailment-like matching operations into a new problem-solving method, to help bridge the gap between the question English and the clean knowledge structures provided by the SME.

References

Clark, P., Harrison, P. Recognizing Textual Entailment with Logical Inference. In Proceedings of 2008 Text Analysis Conference (TAC'08), Gaithsburg, Maryland, 2008.

Giampiccolo, D., Dang, H., Magnin, B., Dagan, I., Cabrio, E., Dolan, B. The Fourth PASCAL Recognizing Textual Entailment Challenge. In Proceedings of 2008 Text Analysis Conference (TAC'08), Gaithsburg, Maryland, 2008.

MacCartney, B., Grenager, T., de Marneffe, M-C., Cer, D., Manning, C. 2006. Learning to recognize features of valid textual entailments. In HLT-NAACL'06, pp 41-48.

Nairn, R., Condoravdi, C., Karttunen, L. Computing relative polarity for textual inference. Inference in Computational Semantics (ICoS-5); 2006 April 20-21; Buxton; UK. http://www.parc.com/publication/1637/computing-relative-polarity-for-textual-inference.html