A Detail-Oriented Approach to Recognizing Textual Entailment (RTE); Negations, Modals, and Temporal Persistence

The Problem:

Textual entailment asks the question “does the truth of some text, \( T \), imply the truth of some other text, \( H \)?” More formally, given a pair of texts \((T, H)\) (also known as the “text” and the “hypothesis”) for which \( T \) is true, may we conclude that \( H \) is true, false, or neither? For example, if \( T = \) “Smith murdered Johnson” and \( H = \) “Johnson is dead,” then the entailment relation is true. If, instead, \( H = \) “Johnson is alive and well,” then the relation is false. And if \( H \) were the statement “Johnson was a helicopter pilot,” then the relation is neither true nor false.

It is important to note that the PASCAL Recognising Textual Entailment Competition (RTE) defines textual entailment as applying not only when \( T \) logically implies \( H \), but also when “a human reading \( T \) would infer that \( H \) is most likely true.” [1]

Textual Entailment is a problem with many potential useful applications in Natural Language Processing. In question answering (QA), textual entailment may be used to determine if a document (or set of documents) implies the truth of a question (or rather, the statement form of a yes-no question). In machine translation (MT), textual entailment could be used for evaluation, testing whether output translations and their corresponding reference translations entail one another. Similarly, textual entailment may be used for automatic evaluation of paraphrasing systems, or document summary systems.

Why Recognizing Textual Entailment is Difficult:

Detecting textual entailment is a difficult task in that there are a number of ways in which entailment relations may occur. Zaenen, et. al. establish an ontology of four different types of textual entailment. [2]

The first, simple entailment, occurs because of strict logical relationships between word meanings, e.g. “Jane planted a willow” entails “Jane planted a tree” in just this manner, as “willow” is defined as a type of tree.

Another kind of entailment is conventional implicature, in which the hypothesis is upheld through the author’s framing of information in the text. For instance, “Amanda denied that the goods were stolen” implies the hypothesis “The goods were stolen” because the author commits to that fact through the use of the word “denied.”

Conversational implicature, on the other hand, implies that the author will share all the information he or she has, e.g. “Mark ate three burgers” entails “Mark did not eat four burgers” under the assumption of conversational implicature.
The fourth, and perhaps most problematic, variety of entailment is that which requires world knowledge. While it is true that the statement “I will attend the next summer Olympics” implies “I will go to Brazil in 2016,” such an inference relies on the knowledge that the next summer Olympics will be held in Brazil in 2016. The degree to which we should require RTE systems to reason with background knowledge is still an open question.

**What Approaches Have Been Tried:**

A variety of approaches have been implemented in attempts at tackling the textual entailment problem. The most basic approaches employ simple similarity testing (e.g. bag-of-words with features like stemming, lemmatization, and part-of-speech tagging). Other approaches represent the texts structurally (e.g. as a graph), and estimate entailment through transforming the structures into one another, and computing a cost function on the transformations. [3] Some approaches frame the problem as proof solving, deriving a “proof” of the hypothesis through iterated rule application. The latter-most approach is a general framework that potentially allows for interfacing with arbitrary modules that contribute more background knowledge or linguistic sophistication. [4]

**What Approaches I Will Try:**

Most of the approaches above are designed with the aim of being general-purpose RTE solvers. As a result, they employ broad statistical approaches (e.g. bag-of-words, n-gram similarity, WordNet distance, etc) without much attention to nuanced linguistic or logical structure. For this project, I intend to focus on one of the many nuanced aspects of the problem that tend to be overlooked by typical bake-off style solvers. Below are a few possible directions for this project.

Negation is one feature that has a complex influence on the entailment relationships depending on the linguistic context. [5] For example, “Jane did not plant a willow” does not entail “Jane did not plant a tree,” but the reverse entailment is true. So in the case of such logical subset relations, negation flips the entailment. In some instances of conventional implicature, negation has no impact on the entailment relation, e.g. “Amanda denied that the goods were stolen” and “Amanda did not deny that the goods were stolen” both entail “The goods were stolen.” Negation also behaves interestingly in colloquial use of intensifiers, e.g. “The food is not terrible” entails “The food is acceptable” or “The food is OK,” but it does not entail its antonym, i.e. “The food is great.” Investigating how negation affects entailment and how we can use this information to improve RTE systems is one possible direction for this project.

Modals are another feature of language that affect entailment relations. The words “might,” “could,” and “should” place verb phrases in the scope of probabilistic, hypothetical, and normative frameworks (respectively) that do not preserve direct the truthfulness of subordinate verb phrase.

Another subtle issue in RTE is that of temporal persistence. When we read a text in the past tense, our expectation of persistence depends largely on the other
content in that text. For example, it is more reasonable to accept that “I moved to Florida last year” entails “I live in Florida now,” than it is to accept that “It rained last week” entails “It is raining right now.” Undoubtedly, this is a problem that involves, to varying degrees, background knowledge. It would be interesting, however, to see if any simple linguistic features correlate strongly with types of statements that exhibit temporal persistence (or a lack thereof), or to what extent such background knowledge can be grouped and represented efficiently.

For this project, I will explore one of the above issues in greater detail and attempt to deal with it in a more focused way than other RTE systems have in the past. A number of RTE solvers are freely available for download online. If possible, I will attempt to incorporate the module(s) I design addressing one of the above issues into one or some of these systems to see if improvements can be made in test performance. How the modules get incorporated will largely depend on how receptive the available systems are to modification. At the very least, they may be used as a sort of black box. Alternatively, the modules I design may be added to a baseline system to see if they provide any benefit.

**Deliverables:**

At the end of the semester, I plan to submit:

- All code, with relevant instructions/comments
- A code demo, if applicable
- A written report, explaining the intellectual content of the project, as well as the results of the project, and an analysis of the results

**Sources:**


[2] Annie Zaenen, Lauri Karttunen, Richard Crouch. Local textual inference: Can it be defined or circumscribed?

