Identifying linguistic knowledge for textual inference

Stacey Bailey
Department of Linguistics
The Ohio State University
Columbus, Ohio, 43210, USA
bailey.462@osu.edu

Abstract

The ability to recognize whether one text can be inferred from another is a critical element of various computational linguistics tasks. The current work explores what kind of linguistic information is useful, or even necessary, in making such inference judgments and how well an inference model using minimal representation and linguistic information performs on an inference judgment task.

1 Introduction

The challenge of recognizing inference is one of deciding what conclusions can be drawn from a given text and of developing a text understanding system that can draw those conclusions automatically. The relevance of inference recognition in a text understanding system cannot be understated. Inferencing is often critical to identifying key concepts, related vocabulary, document topics, etc. in a given task. And there are a wide variety of tasks that have incorporated some inferencing mechanisms, including question answering, machine translation, and text summarization.

For instance, a question answering system that employs inferencing might take a question such as *Who is the author of "Atlas Shrugged"?* and convert it to the form, *<ANSWER> is the author of "Atlas Shrugged"*. In answer retrieval, if the system encounters *Ayn Rand wrote "Atlas Shrugged"*, it will identify the right answer if determines that *Ayn Rand wrote "Atlas Shrugged"* entails *Ayn Rand is the author of "Atlas Shrugged"*.

But many valid inferences are not strictly entailments as is the previous example. For instance, from the sentence *President Reagan was wounded during an assassination attempt*, a reasonable conclusion is that *President Reagan was targeted for assassination*. However, it is not necessarily the case that Reagan was the intended target. Thus, there is no entailment. Nevertheless, the goal of a general inference mechanism is to be able to identify such valid conclusions even when the entailment relation does not hold.

2 Relevant Research

Previous approaches to implementing such inference mechanisms rely largely on surface-level analysis – light semantics and recognizing patterns of sentence entailment. For instance, Ravichandran and Hovy (2002) discuss learning surface patterns useful for question answering. Their patterns are simple regular expressions and reflect the kinds of potential answers expected. For example, the pattern *<NAME> was born in <BIRTHDAY>*, where <NAME> and <BIRTHDAY> are semantic types and the remaining pattern is a surface string, matches the question *When was <NAME> born?*

Monz and de Rijke (2001) also rely on minimal processing for making entailment judgments between sentences, but they do not focus on identifying a set of entailment patterns. Instead, they use an information retrieval notion of text familiarity (inverse document frequency) as the basis of an entailment scoring system.

Dagan and Glickman (2004) also go beyond surface strings to rely on shallow lexical-syntactic representations to build their model of semantic inference recognition, but they, as Ravichandran and Hovy (2002), focus on sentence entailment patterns. For their model, a pattern consists a pair of sentences, syntactic analyses and the direction of entailment between them.

Similarly, the entailment patterns used by Lin and Pantel (2001) are also based on syntactic analysis. Their unsupervised learning algorithm is designed to identify inference rules based on similarity of paths through dependency trees.

3 Current Approach

As suggested above, many of the previous approaches to inference recognition focuses on recognizing entailment between sentences. However, as the Reagan assassination attempt example suggests, conclusions that are more than likely, but not strict entailments, require an inference recognition model that goes beyond patterns of sentence entailment.

In April 2005, the PASCAL Group will host a workshop: Recognising Textual Entailment (RTE), organized by Ido Dagan, Oren Glickman, and Bernardo Magnini. The materials used in the current work are drawn from that challenge and include two development sets of text fragments called text-hypothesis pairs, along with a judgment (TRUE or FALSE) as to whether the text entails the hypothesis. For example, a pair from the first development set is as follows:

<pair id="13" value="TRUE" task="IR">
<t>iTunes software has seen strong sales in Europe.</t>
<h>Strong sales for iTunes in Europe.</h>
</pair>

By entailment, the organizers of the challenge actually mean the more flexible notion of inference of the type described above. Thus, the development and test sets are ideal for the present work.

To begin exploring the boundaries of tools and technology required to build such an inference recognition system, the current paper describes two lines of research. First, what linguistics information is required in order to be able to determine if one text may be inferred from another? The pairs in the RTE development set were analyzed to determine what types of linguistic knowledge might be necessary to make an inference judgment. It was found that while the entailment relation might not hold between the text and hypothesis as a whole, entailment relations often held between subparts of the text and hypothesis. Both lexical entailment (ex: *fell* → *dropped*) and entailment between syntactic constructions (ex: *Prime Minister Silvio Berlusconi* → *Silvio Berlusconi, prime minister,*) are quite common. It is argued that recognition of entailment relations between subparts of text-hypothesis pairs can be used to make a more general inference decision about the pair as a whole.

Second, how well does a system perform that includes minimal linguistic information? A "bag of words" approach is described here that relies on comparing the term overlap between texts and hypothesis pairs. This is compared to Monz and de Rijke's (2001) "bag of words" approach, which requires minimal linguistic information but relies on having a set of related documents in order to calculate inverse document frequency scores. The pairs in the current data set come from a variety of sources and unrelated topics. The current "bag of words" model relies only on the word overlap within pairs, rather than across pairs as Monz and de Rijke's model requires.

References

Ido Dagan and Oren Glickman. 2004. Probabilistic textual entailment: Generic applied modeling of language variability. *Learning Methods for Text Understanding and Mining*, January 26–29, 2004, Grenoble, France.

Oren Glickman and Ido Dagan. 2003. Identifying Lexical Paraphrases From a Single Corpus: A Case Study for Verbs. In Proceedings of *Recent Advantages in Natural Language Processing (RANLP '03)*, September 10-12, 2003.

Dekang Lin and Patrick Pantel. 2001. Discovery of inference rules for question answering. *Natural Language Engineering,* 7(4): 343-360.

Christof Monz and Maarten de Rijke. 2001. Lightweight entailment checking for computational semantics. In *Proc. of the Third Workshop on Inference in Computational Semantics* (IcoS-3).

Deepak Ravichandran and Eduard Hovy. 2002. Learning surface text patterns for a question answering system. In *Proceedings of the 40th ACL conference. Philadelphia.*