Natural Language Understanding and Assessment

Vasile Rus
The University of Memphis, USA

Philip M. McCarthy
The University of Memphis, USA

Danielle S. McNamara
The University of Memphis, USA

Arthur C. Graesser
The University of Memphis, USA

INTRODUCTION

Natural language understanding and assessment is a subset of natural language processing (NLP). The primary purpose of natural language understanding algorithms is to convert written or spoken human language into representations that can be manipulated by computer programs. Complex learning environments such as intelligent tutoring systems (ITSs) often depend on natural language understanding for fast and accurate interpretation of human language so that the system can respond intelligently in natural language. These ITSs function by interpreting the meaning of student input, assessing the extent to which it manifests learning, and generating suitable feedback to the learner. To operate effectively, systems need to be fast enough to operate in the real time environments of ITSs. Delays in feedback caused by computational processing run the risk of frustrating the user and leading to lower engagement with the system. At the same time, the accuracy of assessing student input is critical because inaccurate feedback can potentially compromise learning and lower the student's motivation and metacognitive awareness of the learning goals of the system (Millis et al., 2007). As such, student input in ITSs requires an assessment approach that is fast enough to operate in real time but accurate enough to provide appropriate evaluation.

One of the ways in which ITSs with natural language understanding verify student input is through *matching*. In some cases, the match is between the user input and a pre-selected *stored answer to a question*, *solution to a problem*, *misconception*, or other form of *benchmark response*. In other cases, the system evaluates the degree to which the student input varies from a complex representation or a dynamically computed structure. The computation of matches and similarity metrics are limited by the fidelity and flexibility of the computational linguistics modules.

The major challenge with assessing natural language input is that it is relatively unconstrained and rarely follows brittle rules in its computation of spelling, syntax, and semantics (McCarthy et al., 2007). Researchers who have developed tutorial dialogue systems in natural language have explored the accuracy of matching students' written input to targeted knowledge. Examples of these systems are AutoTutor and Why-Atlas, which tutor students on Newtonian physics (Graesser, Olney, Haynes, & Chipman, 2005; VanLehn , Graesser, et al., 2007), and the iSTART system, which helps students read text at deeper levels (McNamara, Levinstein, & Boonthum, 2004). Systems such as these have typically relied on statistical representations, such as latent semantic analysis (LSA; Landauer, McNamara, Dennis, & Kintsch, 2007) and content word overlap metrics (McNamara, Boonthum, et al., 2007). Indeed, such statistical and word overlap algorithms can boast much success. However, over short dialogue exchanges (such as those in ITSs), the accuracy of interpretation can be seriously compromised without a deeper level of lexico-syntactic textual assessment (McCarthy et al., 2007). Such a lexico-syntactic approach, *entailment evaluation*, is presented in this chapter. The approach incorporates deeper natural language processing solutions for ITSs with natural language exchanges while remaining sufficiently fast to provide real time assessment of user input.

BACKGROUND

Entailment evaluations help in the assessment of the appropriateness of student responses during ITS exchanges. Entailment can be distinguished from three similar terms (implicature, paraphrase, and elaboration), all of which are also important for assessment in ITS environments (McCarthy et al, 2007).

The terms *entailment* is often associated with the highly similar concept of *implicature*. The distinction is that entailment is reserved for linguistic-based inferences that are closely tied to *explicit* words, syntactic constructions, and formal semantics, as opposed to the knowledge-based *implied* referents and references, for which the term implicature is more appropriate (McCarthy et al., 2007). Implicature corresponds to the controlled knowledge-based elaborative inferences defined by Kintsch (1993) or to knowledge-based inferences defined in the inference taxonomies in discourse psychology (Graesser, Singer, & Trabasso, 1994).

The terms *paraphrase* and *elaboration* also need to be distinguished from entailment. A *paraphrase* is a reasonable restatement of the text. Thus, a paraphrase is a form of entailment, yet an entailment is not necessarily a paraphrase. This asymmetric relation can be understood if we consider that *John went to the store* is entailed by (but not a paraphrase of) *John drove to the store to buy supplies*. The term *elaboration* refers to information that is generated inferentially or associatively in response to the text being analyzed, but without the systematic and sometimes formal constraints of entailment, implicature, or paraphrase. Examples of each term are provided below for the sentence *John drove to the store to buy supplies*.

Entailment: *John went to the store.*
(Explicit, logical implication based on the text)

Implicature: *John bought some supplies.*
(Implicit, reasonable assumption from the text, although not explicitly stated in the text)

Paraphrase: *He took his car to the store to get things that he wanted.*
(Reasonable re-statement of all and only the critical information in the text)

Elaboration: *He could have borrowed stuff.*
(Reasonable *reaction* to the text)

Evaluating entailment is generally referred to as the task of *recognizing textual entailment* (RTE; Dagan, Glickman, & Magnini, 2005). Specifically, it is the task of deciding, given two text fragments, whether the meaning of one text logically infers the other. When it does, the evaluation is deemed as T (the entailing text) entails H (the entailed hypothesis). For example, a text (from the RTE data) of *Eyeing the huge market potential, currently led by Google, Yahoo took over search company Overture Services Inc last year* would entail a hypothesis of *Yahoo bought Overture*. The task of recognizing entailment is relevant to a large number of applications, including machine translation, question answering, and information retrieval.

The task of textual entailment has been a priority in investigations of information retrieval (Monz & de Rijke, 2001) and automated language processing (Pazienza, Pennacchiotti, & Zanzotto, 2005). In related work, Moldovan and Rus (2001) analyzed how to use unification and matching to address the *answer correctness* problem. Similar to entailment, *answer correctness* is the task of deciding whether candidate answers logically imply an ideal answer to a question.

THE LEXICO-SYNTACTIC ENTAILMENT APPROACH

A complete solution to the textual entailment challenge requires linguistic information, reasoning, and world knowledge (Rus, McCarthy, McNamara, & Graesser, in press). This chapter focuses on the role of linguistic information in making entailment decisions. The overall goal is to produce a light (i.e. computationally inexpensive), but accurate solution that could be used in interactive systems such as ITSs. Solutions that rely on processing-intensive deep representations (e.g., frame semantics and reasoning) and large structured repositories of information (e.g., ResearchCyc) are impractical for interactive tasks because they result in lengthy response times, causing user dissatisfaction.

One solution for recognizing textual entailment is based on subsumption. In general, an object X subsumes