Chapter 15
Applying NLP Metrics to Students’ Self-Explanations

G. Tanner Jackson
Arizona State University, USA

Danielle S. McNamara
Arizona State University, USA

ABSTRACT

Intelligent Tutoring Systems (ITSs) are becoming an increasingly common method for students to engage with and learn course material. ITSs are designed to provide students with one-on-one learning that is tailored to their own pace and needs. These systems can adapt to each users’ individual knowledge and ability level to provide the most pedagogically effective learning environment. Tutoring systems have been designed that cover a variety of topics, including both well-defined and ill-defined domains. ITSs have seen great success within well-defined domains, where the topic itself provides only a limited set of responses. For example, in the domain of algebra, there is a limited set of possible actions that can be performed to solve for an unknown variable. Knowing this complete set of actions allows the tutoring system to predict all possible responses from the user. In contrast, ill-defined domains are more abstract and open ended. Reading comprehension is an ill-defined, open ended domain that can incorporate text from any subject, and involve numerous processes and problems for the learner. The number of associations that learners can make with a given text (e.g., based on personal memories, previous courses, ideas within different parts of the same text, etc.) is virtually infinite. These associations make it almost impossible to predict how a user will respond to a text. In addition to working with more abstract concepts, ITSs within ill-defined domains often have the added challenge of interpreting natural language user input. Incorporating natural language allows learners to use their own words and ideas as they interact with the content; however, this also increases the ambiguity of the interaction and decreases the system’s ability to build a precise model of the learner. Building an accurate learner model is essential for the system to adapt the interaction in a pedagogically appropriate manner.

DOI: 10.4018/978-1-60960-741-8.ch015
INTRODUCTION

Recent advances in Natural Language Processing (NLP) have dramatically increased the ability to interpret user language input. These advances have made it possible to develop ITSs that rely heavily on natural language input to drive the interaction with the system (Graesser, Lu, Jackson, Mitchell, Ventura, Olney, & Louwerse, 2004; McNamara, Boonthum, Levinstein, & Millis, 2007; VanLehn, Graesser, Jackson, Jordan, Olney, & Rose, 2007). An ITS called Interactive Strategy Training for Active Reading and Thinking (iSTART) is one of these systems that relies on users’ natural language. iSTART utilizes a complex NLP algorithm to evaluate student contributions and select an appropriate system response. Previous evaluations of the algorithm have provided strong evidence that it produces an adequate representation of student input (Jackson, Guess, McNamara, 2010; McNamara, Boonthum, Levinstein, & Millis, 2007). Nonetheless, there are still potential improvements that could increase the assessment accuracy. The aim of the current chapter is to assess the viability of including new linguistic measures from Coh-Metrix (Graesser, McNamara, Louwerse, & Cai, 2004) to the existing iSTART assessment algorithm.

iSTART

iSTART is a web-based tutoring system designed to improve high school and college students’ reading comprehension by providing instruction and practice using self-explanation and reading strategies. The iSTART system was originally modeled after a human-based intervention called Self-Explanation Reading Training, or SERT (McNamara, 2004; McNamara & Scott, 1999; O’Reilly, Best, & McNamara, 2004). The automated iSTART system has consistently produced gains equivalent to the human-based SERT program (Magliano et al., 2004; O’Reilly, Sinclair, & McNamara, 2004; O’Reilly, Best, & McNamara, 2004). Unlike SERT, iSTART is web-based, and can potentially provide training to schools or individuals with internet access. Furthermore, because it is automated, it can work with students on an individual level and provide self-paced instruction. iSTART also maintains a record of student performance and can use this information to adapt its feedback and instruction for each student. Lastly, iSTART combines pedagogical agents and automated linguistic analysis to engage the student in an interactive dialogue and create an active learning environment (e.g., Bransford, Brown, & Cocking, 2000; Graesser, Hu, & Person, 2001; Graesser, Hu, & McNamara, 2005; Louwerse, Graesser, & Olney, 2002).

iSTART Modules

iSTART incorporates pedagogical agents that engage users with the system and tutor them on how to correctly apply various reading strategies. The agents were designed to introduce students to the concept of self-explanation and to demonstrate specific strategies that could potentially enhance their ability to self-explain and in turn, their reading comprehension. The iSTART program consists of three system modules that implement the pedagogical principle of modeling-scaffolding-fading: introduction, demonstration, and practice.

The introduction module uses a classroom-like discussion format between three animated agents (a teacher, Dr. Julie, and two student agents, Sheila and Mike) to present the relevant reading strategies within iSTART. These agents interact with each other, providing students with information, posing questions to each other, and providing example self-explanations to illustrate appropriate strategy use (including counterexamples). These interactions exemplify the active processing that students should use when generating their own self-explanations. For each strategy, the students also answer multiple-choice questions.
Related Content

Hidden Markov Model Based Visemes Recognition, Part I: AdaBoost Approach
[www.igi-global.com/chapter/hidden-markov-model-based-visemes/31073?camid=4v1a](www.igi-global.com/chapter/hidden-markov-model-based-visemes/31073?camid=4v1a)

Parsing Bangla Grammar Using Context Free Grammar
[www.igi-global.com/chapter/parsing-bangla-grammar-using-context-free-grammar/108758?camid=4v1a](www.igi-global.com/chapter/parsing-bangla-grammar-using-context-free-grammar/108758?camid=4v1a)

Semantic Methods for Textual Entailment
Andrew Neel and Max H. Garzon (2012). *Applied Natural Language Processing: Identification, Investigation and Resolution* (pp. 479-494).
[www.igi-global.com/chapter/semantic-methods-textual-entailment/61066?camid=4v1a](www.igi-global.com/chapter/semantic-methods-textual-entailment/61066?camid=4v1a)

Sentiment Classification: Facebook’ Statuses Mining in the “Arabic Spring” Era
[www.igi-global.com/chapter/sentiment-classification/133873?camid=4v1a](www.igi-global.com/chapter/sentiment-classification/133873?camid=4v1a)