NLP Techniques in Intelligent Tutoring Systems

Chutima Boonthum
Hampton University, USA

Irwin B. Levinstein
Old Dominion University, USA

Danielle S. McNamara
University of Memphis, USA

Joseph P. Magliano
University of Memphis, USA

Keith K. Millis
University of Memphis, USA

INTRODUCTION

Many Intelligent Tutoring Systems (ITSs) aim to help students become better readers. The computational challenges involved are (1) to assess the students' natural language inputs and (2) to provide appropriate feedback and guide students through the ITS curriculum. To overcome both challenges, the following non-structural Natural Language Processing (NLP) techniques have been explored and the first two are already in use: word-matching (WM), latent semantic analysis (LSA, Landauer, Foltz, & Laham, 1998), and topic models (TM, Griffiths & Steyvers, 2007).

This article describes these NLP techniques, the iSTART (Strategy Trainer for Active Reading and Thinking, McNamara, Levinstein, & Boonthum, 2004) intelligent tutor and the related Reading Strategies Assessment Tool (R-SAT, Magliano *et al.*, 2006), and how these NLP techniques can be used in assessing students' input in iSTART and R-SAT. This article also discusses other related NLP techniques which are used in other applications and may be of use in the assessment tools or intelligent tutoring systems.

BACKGROUND

Interpreting text is critical for intelligent tutoring systems (ITSs) that are designed to interact meaningfully with, and adapt to, the users' input. Different ITSs use different Natural Language Processing (NLP) techniques in their system. NLP systems may be structural, *i.e.,* focused on grammar and logic, or non-structural, *i.e.,* focused on words and statistics. This article deals with the latter.

Examples of the structural approach include ExtrAns (Extracting Answers from technical texts question-answering system; Molla *et al.*, 2003) which uses minimal logical forms (MLF; that is, the form of first order predicates) to represent both texts and questions and C-Rater (Leacock & Chodorow, 2003) which scores short-answer questions by analyzing the conceptual information of an answer in respect to the given question. Turning to the non-structural approach, AutoTutor (Graesser *et al.*, 2000) uses LSA to analyze the student's input against expected sets of answers and CIRCSIM-Tutor (Kim *et al.*, 1989) uses a word-matching technique to evaluate students' short answers. The systems considered more fully below, iSTART (McNamara *et al.*, 2004) and R-SAT (Magliano *et al.*, 2006) use both word-matching and LSA in assessing quality of students' self-explanation. Topic models (TM) were explored in both systems, but have not yet been integrated.

MAIN FOCUS OF THE CHAPTER

This article presents three non-structural NLP techniques (WM, LSA, and TM) which are currently used

or being explored in reading strategies assessment and training applications, particularly, iSTART and R-SAT.

Word Matching

Word matching is a simple and intuitive way to estimate the nature of an explanation. There are two ways to compare words from the reader's input (either answers or explanations) against benchmarks (collections of words that represent a unit of text or an ideal answer): (1) Literal word matching and (2) Soundex matching.

Literal word matching – Words are compared character by character and if there is a match of sufficient length then we call this a *literal match*. An alternative is to count words that have the same stem (*e.g.,* indexer and indexing) as matching. If a word is short a complete match may be required to reduce the number of false-positives.

Soundex matching – This algorithm compensates for misspellings by mapping similar characters to the same soundex symbol (Christian, 1998). Words are transformed to their soundex code by retaining the first character, dropping the vowels, and then converting other characters into soundex symbols: *1* for *b, p; 2* for *f, v; 3* for *c, k, s; etc.* Sometimes only one consecutive occurrence of the same symbol is retained. There are many variants of this algorithm designed to reduce the number of false positives (*e.g.,* Philips, 1990). As in literal matching, short words may require a full soundex match while for longer words the first *n* soundex symbols may suffice.

Word-matching is also used in other applications, such as, CIRCSIM-Tutor (Kim *et al.*, 1989) on short-answer questions and Short Essay Grading System (Ventura, 2004) on questions with ideal expert answers.

Latent Semantic Analysis (LSA)

Latent Semantic Analysis (LSA; Landauer, Foltz, & Laham, 1998) uses statistical computation to extract and represent the meaning of words. Meanings are represented in terms of their similarity to other words in a large corpus of documents. LSA begins by finding the frequency of terms used and the number of co-occurrences in each document throughout the corpus and then uses a powerful mathematical transformation to find deeper meanings and relations between words.

When measuring the similarity between text-objects, LSA's accuracy improves with the size of the objects, so it provides the most benefit in finding similarity between two documents but as it does not take word order into account, short documents may not receive the full benefit. The details for constructing an LSA corpus matrix are in Landauer & Dumais (1997). Briefly, the steps are: (1) select a corpus; (2) create a term-document-frequency (TDF) matrix; (3) apply Singular Value Decomposition (SVD; Press *et al.*, 1986) to the TDF matrix to decompose it into three matrices (L x S x R; where S is a scaling, matrix). The leftmost matrix (L) becomes the LSA matrix of that corpus. The optimal size is usually in the range of 300–400 dimensions. Hence, the LSA matrix dimensions become N x D where N is the number of unique words in the entire corpus and D is the optimal dimension (reduced from the total number of documents in the entire corpus).

The similarity of terms (or words) is computed by comparing two rows, each representing a term vector. This is done by taking the cosine of the two term vectors. To find the similarity of sentences or documents, (1) for each document, create a document vector using the sum of the term vectors of all the terms appearing in the document and (2) calculate a cosine between two document vectors. Cosine values range from ±1 where +1 means highly similar.

To use LSA in the tutoring systems, a set of benchmarks are created and compared with the trainee's input. Examples benchmarks are the current target sentence, previous sentences, and the ideal answer. A high cosine value between the current sentence benchmark and the reader's input would indicate that the reader understood the sentence and was able to paraphrase what was read. To provide appropriate feedback, a number of cosines are computed (one for each benchmark). Various statistical methods, such as discriminant analysis and regression analysis, are used to construct the feedback formula. McNamara *et al.* (2007) describe various ways that LSA can be used to evaluate the reader's explanations: either LSA alone or a combination of LSA with WM. The final conclusion is that a fully-automated (*i.e.*, less hand-crafted benchmarks construction), combined system produces the better results.

There are a number of other intelligent tutoring systems that use LSA in their feedback system, for examples, Summary Street (Steinhart, 2001), Auto-

Tutor (Greasser *et al.*, 2000), and Tutoring System (Lemaire, 1999).

Topic Models

The Topic Models approach (TM; Steyvers & Griffiths, 2007) applies a probabilistic model to find a relationship between terms and documents in terms of topics. A document is considered to be generated probabilistically from a number of topics where each topic consists of a number of terms, each given a probability of selection if that topic is used. By using a TM matrix, the probability that a certain topic was used in the creation of a given document is estimated. If two documents are similar, the estimates of the topics within these documents should be similar. TM is similar to LSA, except that a term-document frequency matrix is factored into two matrices instead of three: one is the probabilities of terms belonging to the topics (the TM matrix), the other the probabilities of topics belonging to the documents. The Topic Modeling Toolbox (Steyvers & Griffiths, 2007) can be used to construct a TM matrix,

To measure the similarity between documents, the Kullback Leibler distance (KL-distance: Steyvers & Griffiths, 2007) is recommended, rather than the cosine measure (which can also be used). Using TM in a tutoring system is similar to using LSA, where a set of benchmarks is defined and the reader's input is compared against each benchmark. The only different is the use of KL-distance instead of LSA-cosine value. The preliminary results of investigating TM in place of LSA (Boonthum, Levinstein, & McNamara, 2006) indicate that TM is as good as LSA alone (correlation between computerized-scores and human rating scores), but a little bit lower than a combined system using both WM and LSA. This suggests that the TM should be further investigated in combination with WM or LSA or both.

TM is mostly used in document clustering (grouping documents based on relevancy or similar topics; Buntine *et al.*, 2005), data mining (Tuulos & Tirri, 2004), and search engines (Perkiö *et al.*, 2004). A variation on TM by Steyvers & Griffiths (2007), is *Probabilistic Latent Semantic Analysis* (PLSA; Hofmann, 2001) which models each document as generated from a number of hidden topics and each topic has its features defined as the conditional probabilities of word occurrences in that topic.

iSTART and RSAT applications

iSTART (Interactive Strategy Trainer for Active Reading and Thinking) is a web-based, automated tutor designed to help students become better readers using multi-media technology. It provides adolescent to college-aged students with a program of self-explanation and reading strategy training (McNamara *et al.*, 2004) called Self-Explanation Reading Training, or SERT (see McNamara *et al.*, 2004). iSTART consists of three modules: Introduction (description of SERT and reading strategies), Demonstration (illustration of how these reading strategies can be used), and Practice (hands-on practice of these reading strategies). In the Practice module, students practice using reading strategies by typing self-explanations of sentences. The system evaluates each explanation and then provides appropriate feedback to the student. If the explanation is irrelevant or too short compared to the given sentence and passage, the student is required to add more information. Otherwise, the feedback is based on the level of its overall quality.

The computational challenge is to provide appropriate feedback to the students about their explanations. Doing so requires capturing some sense of both the meaning and quality of their explanation. A combination of word-matching and LSA provided better results (comparing the computerized-score using NLP techniques to the human rating score and having higher correlation between these two sets of scores) than either separately (McNamara, Boonthum, Levinstein, & Millis, 2007).

R-SAT (Reading Strategy Assessment Tool; Maglino *et al.*, 2007) is an automated web-based reading assessment tool designed to measure readers' comprehension and spontaneous use of reading strategies. The R-SAT is similar to the iSTART Practice module in the sense that it presents passages to the reader one sentence at a time and asks for the reader's input. The difference is that, instead of an explanation, R-SAT asks either an indirect ("What are your thoughts regarding your understanding of the sentence in the context of the passage?") or a direct question (*e.g.,* Why did the miller want to marry the girl?") at pre-selected target sentences. The answers to the indirect questions are evaluated on how they are related to the given sentence and passage; the answers to the direct questions are assessed by comparing them to ideal answers.

The problem is to analyze the answers and generate a set of scores for overall comprehension and strategy usage. Ultimately, these scores can be used as a pre-assessment for iSTART allowing the trainer to individualize the iSTART curriculum based on the reader's needs. R-SAT was initially proposed to use word-matching, LSA, and other techniques beyond LSA. However, during the course of development, word-matching was found to produce better results than LSA or in combination with LSA.

FUTURE TRENDS

These three NLP techniques (WM, LSA, and TM) are used in the ongoing research on assessing and improving comprehension skills via reading strategies in the R-SAT and iSTART projects. WM and LSA have been extensively investigated for iSTART and to some extent in R-SAT. The lack of success of LSA compared to the simpler WM in R-SAT is somewhat surprising and may be due to particular features of the algorithms used or to the variety of text genres used in R-SAT. Future work is planned with modified algorithms and substituting genre-specific LSA spaces for the general space now used. In addition TM needs further exploration, especially in its use with small units of text where the recommended Kullback Leibler distance has not proven particularly effective.

CONCLUSION

The purpose of this article is to describe three NLP techniques and how they can be used in assessment tools and intelligent tutoring systems. For iSTART to teach reading strategies effectively, it must be able to deliver valid feedback on the quality of the explanations that a reader produces and therefore the system must understand, at least to some extent, the explanation. Of course, automating natural language understanding has been extremely challenging, especially for non-restrictive content domains like explaining a freely-entered text t. Algorithms such as LSA open up a number of possibilities to systems such as iSTART: in essence LSA provides a 'simple' algorithm that allowed tutoring systems to provide appropriate feedback to students (see Landauer *et al.*, 2007). The results presented in Boonthum *et al.* (2006) show that the topic model similarly offers a wealth of possibilities in natural language processing. For R-SAT to measure a reader's comprehension and reading skills accurately, like iSTART it must also be able to understand, to some extent, what a reader says, especially when he/she is asked to describe their current thoughts. Although LSA is a good candidate, simple word matching against various benchmarks seems adequate to provide satisfactory results especially when aggregated over several explanations (see Magliano *et al.*, 2006). It is also demonstrates that a combination of techniques produces better results than using one technique on its own.

REFERENCES

Boonthum, C., Levinstein, I.B., & McNamara, D.S. (2006). Evaluating Self-Explanations in iSTART: Word Matching, Latent Semantic Analysis, and Topic Models. In A. Kao & S. Poteet (Eds.), *Text Mining and Natural Language Processing*, Springer. 91-106.

Buntine, W., Löfström, J., Perttu, S., & Valtonen, K. (2005). Topic-Specific Scoring of Documents for Relevant Retrieval. In *Workshop on Learning in Web Search (LWS-2005)*, pp 34-41.

Christian. P. (1998) Soundex – can it be improved? *Computers in Genealogy*, 6 (5).

Graesser, A., Wiemer-Hastings, P., Wiemer-Hastings, K., Harter, D., Person, N., & TRG. (2000). Using Latent Semantic Analysis to evaluate the contributions of students in AutoTutor. *Interactive Learning Environments*, 8 , 149-169.

Hofmann, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning*, pp 177-196.

Kim, N., Evens, M.W., Michael, J.A., & Rovick, A.A. (1989) CIRCSIM-Tutor: An Intelligent Tutoring System for Circulatory Physiology. In Maurer, H. (ed.), *Computer-Assisted Learning: 2nd International Conference* (ICCAL-89), pp. 254-266. Berlin: Springer-Verlag.

Landauer, T.K., Foltz, P.W., & Laham, D. (1998). Intorduction to Latent Semantic Analysis. *Discourse Processes*, 25, 259-284.

Landauer, T., McNamara, D.S., Dennis, S., & Kintsch, W. (2007). *A Handbook of Latent Semantic Analysis* Mahwah, NJ: Erlbaum.

Leacock, C., & Chodorow, M. (2003). C-rater: Automated scoring of short-answer questions. *Computers and the Humanities*, 37(4), 389-405.

Lemaire, B. (1999). Tutoring Systems Based on Latent Semantic Analysis. In *Artificial Intelligence in Education* (AIED-99), S. Lajoie and M. Vivet (eds.), IOS Press, Amsterdam, pp. 527-534.

Magliano, J.P., Millis, K.K., Gilliam, S., Levinstein, I.B., & Boonthum, C. (2006). Assessing Reading Comprehension with Verbal Protocols and Latent Semantic Analysis. In *the Proceeding of the 47th Annual Meeting of the Psychonomic Society*, Houston, TX.

McNamara, D.S., Boonthum, C., Levinstein, I.B., & Millis, K.K. (2007). Using LSA and word-based measures to assess self-explanations in iSTART. In T. Landauer *et al.* (Eds.), *A Handbook of Latent Semantic Analysis* Mahwah, NJ: Erlbaum.

McNamara, D.S., Levinstein, I.B., & Boonthum, C. (2004). iSTART: Interactive Strategy Trainer for Active Reading and Thinking. Submitted to *Behavioral Research Methods, Instruments, and Computers*, 36, 222-233.

Molla, D., Schwitter, R., Rinaldi, F., Dowdall, J., & Hess, M. (2003). ExtrAns: Extracting Answers from Technical Texts. *IEEE Intelligent System* 18(4): 12-17.

Perkiö, J., Buntine., W., & Perttu, S. (2004). Exploring Independent Trends in a Topic-Based Search Engine. In *Proceedings of the Web Intelligence Conference (WI-2004)*, pp. 664-668.

Philips, L. (1990). Hanging on the Metaphone. *Computer Language* 7(12).

Steinhart, D. (2001). Summary Street: An intelligent tutoring system for improving student writing through the use of latent semantic analysis. *Ph.D. dissertation*, Dept. Psychology, Univ. Colorado, Boulder.

Steyvers, M., & Griffiths, T. (2007). Probabilistic Topic Models. In T. Landauer, D.S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis*, Mahwah, NJ: Erlbaum.

Tuulos, V.H. & Tirri, H. (2004). Combining Topic Models and Social networks for Chat Data Mining. In *Proceedings of on Web Intelligence Conference (WI-2004)*, pp. 206-213.

Ventura, M.J., Franchescetti, D.R., Pennumatsa, P., Graesser, A.C., Jackson, G.T., Hu, X., Cai, Z., & TRG. (2004). Combining Computational Models of Short Essay Grading for Conceptual Physics Problems. In J.C. Lester *et al.* (Eds.), Intelligent Tutoring Systems (pp. 423-431). Berlin, Germany: Springer.

KEY TERMS

Intelligent Tutoring System (ITS): Also called Intelligence Computer-Aided Instruction (ICAI), a personal training assistant that captures the subject matter and teaching expertise and individualize the curriculum to meet each learner's needs in order to master the subject matter. Its main goal is to provide benefits of the one-on-one instruction: lessons are conducted at the learner's own pace; practices are interactive so the learner can improve their weaker skills; and real-time question answering clarify learner's doubts or misunderstanding; and an individualized curriculum based on the learner's needs.

Kullback Leibler Distance (KL-distance): A natural distance function from a "true" probability distribution to a "target" probability distribution. It can be interpreted as the expected extra message-length per datum due to using a code based on the wrong (target) distribution compared to using a code based on the true distribution.

Latent Semantic Analysis (LSA): A natural language processing technique that analyses relationships between a set of documents and terms within these documents. LSA was created in 1990 for information retrieval and is sometimes called latent semantic indexing (LSI).

LSA Cosine: A measurement of a relation between two vector-units. A unit can be as small as a word or as large as an entire document. It can be computed using the dot-product of two vectors where each vector is a representation of a unit (word, sentence, paragraph, or whole document).

Probabilistic Latent Semantic Analysis (PLSA): A statistical techniques for the analysis of two-mode and co-occurrence data, which has applications in information retrieval and filtering, natural language processing, machine learning from text, and related areas. PLSA evolved from LSA but focuses more on the relationship of topics within documents.

Protocols: Any verbal input that students or readers produce during a session. This can be a set of explanations or answers to direct questions.

Self-Explanation and Reading Strategy Trainer (SERT): Pedagogy uses five strategies to help students become a better reader. The reading strategies include (1) comprehension monitoring, being aware of one's own understanding of the text; (2) paraphrasing, or restating the text in different words; (3) elaboration, using prior knowledge or experiences to understand the text (domain-specific knowledge-based inferences) or using common-sense or logic to understand the text (general knowledge based inferences); (4) predictions, predicting what the text will say next; and (5) bridging, understanding the relation between separate sentences of the text.

Word Matching (WM): A simple way to compare words. Literal match is done by comparing character by character, while Soundex match transforms each word into a Soundex code, similar to phonetic spelling.