Open Research Online

The Learning Grid and E-Assessment using Latent Semantic Analysis

Debra Trusso Haley, Pete Thomas, Bashar Nuseibeh, Josie Taylor, Paul Lefrere
Computing Research Centre, The Open University,
Walton Hall, Milton Keynes, UK MK7 6AA
email: [D.T.Haley, P.G.Thomas, B.Nuseibeh, J.Taylor] [at] open.ac.uk

Abstract. E-assessment is an important component of e-learning and e-qualification. Formative and summative assessment serve different purposes and both types of evaluation are critical to the pedagogical process. While students are studying, practicing, working, or revising, formative assessment provides direction, focus, and guidance. Summative assessment provides the means to evaluate a learner's achievement and communicate that achievement to interested parties. Latent Semantic Analysis (LSA) is a statistical method for inferring meaning from a text. Applications based on LSA exist that provide both summative and formative assessment of a learner's work. However, the huge computational needs are a major problem with this promising technique. This paper explains how LSA works, describes the breadth of existing applications using LSA, explains how LSA is particularly suited to e-assessment, and proposes research to exploit the potential computational power of the Grid to overcome one of LSA's drawbacks.

Keywords: Grid, LSA, latent semantic analysis, e-assessment, assessment

Introduction

This paper describes Latent Semantic Analysis (LSA) and presents a research plan for combining the potential computational power of the Grid with LSA's ability to provide immediate, accurate, personalised, and content-based feedback. This electronic feedback, or e-assessment, is an important component of e-learning and e-qualification.

Types of Assessment

Formative assessment provides direction, focus, and guidance concurrent with the learner engaging in some learning process. E-assessment can provide ample help to a learner without requiring added work by a human tutor. A learner can benefit from private, immediate, and convenient feedback. Summative assessment, on the other hand, happens at the end of a learning episode or activity. It evaluates a learner's achievement and communicates that achievement to interested parties. Summative assessment shares the virtues of formative assessment while improving the ability to achieve more objective grading results than those that can occur when many markers are assessing hundreds of students.

LSA and the Grid

The Grid is often described as the next generation of the Web [1]. The success of the Grid depends on useful applications being available [2]. Dahn [3] suggests that a major educational application of the Grid is as a storage medium for learning resources featuring easy location, retrieval, and sharing by students. E-assessment using LSA is another potential Grid application with major pedagogical benefits to learners. It offers immediate feedback to learners exactly when they can most benefit.

1. What is Latent Semantic Analysis?

1.1. About LSA

Researchers at Bellcore developed LSA, a statistical-based method for inferring meaning from a text. Landauer et. al. [4] give a more formal definition: "Latent Semantic Analysis is a theory and method for extracting and representing the contextual-usage meaning of words by statistical computations applied to a large corpus of text". It was first used as an information retrieval technique [5] in the 1980s. By 1997, Landauer and Dumais [6] asserted that LSA could serve as a model for the human acquisition of knowledge. They developed their theory after creating a mathematical information retrieval tool and observing unexpected results from its use. They claimed that LSA solves Plato's problem, that is, *how do people learn so much when presented with so little?* Their answer, oversimplified but essentially accurate, is the inductive process: LSA "induces global knowledge indirectly from local co-occurrence data in a large body of representative text" [6].

From the original application for retrieving information, the use of LSA has evolved to systems that more fully exploit its ability to extract and represent meaning. Recent applications based on LSA compare a sample text with a pre-existing training corpus to judge the quality of the sample. The corpora are very large; for example, Summary Street, an LSA-based instructional software system, uses a corpus of 11 million words [7].

1.2. How LSA Works

Even the developers of LSA understand that its results can seem magical [6]. However, a thorough understanding of the mathematical and statistical underpinnings of the method can provide some clarity.

To use LSA, researchers amass a suitable corpus of text. (Exactly what corpus is most suitable for which purpose is an issue requiring further research.) They create a term-by-document matrix where the columns are documents and the rows are terms [5]. A term is a subdivision of a document; it can be a word or phrase or some other unit. A document can be a sentence, a paragraph, a textbook, or some other unit. In other words, documents contain terms. The elements of the matrix are weighted word counts of how many times each term appears in each document.

After it creates the matrix, LSA decomposes it into three matrices using Singular Value Decomposition (SVD), a well-known technique [8] that is the general case of

factor analysis, which decomposes a square matrix with like rows and columns [6]. Deerwester et. al., [5] describe the process as follows.

Let t = the number of terms, or rows
d = the number of documents, or columns
X = a t by d matrix

Then, after applying SVD, X = TSD, where

m = the number of dimensions, m <= min(t,d)
T = a t by m matrix
S = an m by m diagonal matrix, i.e., only diagonal entries have non-zero values
D = an m by d matrix

The "magic" performed by LSA is to reduce S, the diagonal matrix created by SVD, to an appropriate number of dimensions resulting in S'. The product of TS'D is the least-squares best fit to X, the original matrix [5].

People often describe LSA as analyzing co-occurring terms when, actually, it does more: Landauer and Dumais [6] explain that the new matrix captures the "latent transitivity relations" among the terms. Terms not appearing in an original document are represented in the new matrix as if they actually were in the original document [6]. LSA's ability to induce transitive meanings is especially important considering that Furnas et. al. [9] report fewer than 20% of paired individuals will use the same term to refer to the same common concept.

LSA exploits what can be named the transitive property of semantic relationships: If $A \rightarrow B$ and $B \rightarrow C$, then $A \rightarrow C$ (where $\rightarrow$ stands for *is semantically related to)*. However, the similarity to the transitive property of equality is not perfect. Two words widely separated in the transitivity chain can have a weaker relationship than closer words. For example, LSA might find that copy $\rightarrow$ duplicate $\rightarrow$ double $\rightarrow$ twin $\rightarrow$ sibling. Copy and duplicate are much closer semantically than copy and sibling.

Finding the correct number of dimensions for the new matrix created by SVD is critical; if it is too small, the structure of the data is not captured. Conversely, if it is too large, sampling error and unimportant details remain, e.g., grammatical variants [5, 7, 8]. Empirical work shows the correct number of dimensions to be about 300 [6, 7].

Creating the matrices using SVD and reducing the number of dimensions, often referred to as training the system, requires a lot a computing power; it can take hours or days to complete the processing [8]. Fortunately, once the training is complete, it takes just seconds for LSA to evaluate a text sample [8].

1.3. Existing Applications using LSA

The earliest application of LSA was Latent Semantic Indexing (LSI) [5, 10]. LSI provided an advantage over keyword-based methods in that it could induce associative meanings of the query [11] rather than relying on exact matches.

Soto [12] suggests another use for LSA: improving the interfaces of software systems. Users more easily learn and remember the function of menu labels when they are related semantically to the users' conceptions. The HCI community uses cognitive walkthroughs, among other things, to assess the learnability and memorability of menu

labels. Soto [12] suggests LSA could be a cheaper, faster replacement of cognitive walkthroughs.

Researchers are carrying out some interesting work into medical uses of LSA. Wu et. al. [13], as referenced in Chung & O'Neil [14], are using LSA to classify protein sequences. Skoyle [15] is using the theory of Landauer and Dumais [6] to investigate whether autism results from a failure in an individual's ability to create meaning by an indirect process - the induction modelled by LSA. Campbell and Pennebaker [16] are using LSA to demonstrate linkages between writing about traumatic events and improving health.

Researchers [17] have achieved good results in matching texts to a learner's reading ability using LSA. If a text is too easy, a learner doesn't learn anything; if a text is too hard, it can be incomprehensible. They refer to the Goldilocks principle [17] of using texts at just the right difficulty level – slightly beyond the learner's ability and knowledge.

Much work is being done in the area of using LSA to grade essays automatically and to provide content-based feedback. One of the great advantages of automatic assessment of essays is its ability to provide helpful, immediate feedback to the student without burdening the teacher. This application is particularly suited to distance education, where opportunities for one-on-one tutoring are infrequent or non-existent [18]. Existing systems include Apex [19], Autotutor [20], Intelligent Essay Assessor [17], Select-a-Kibitzer [8], and Summary Street [7, 18]. They differ in details of audience addressed, subject domain, and advanced training required by the system [8]. They are similar in that they are LSA-based, web-based, and provide the scaffolding, feedback, and unlimited practice opportunities without increasing a teacher's workload [18]. See [8] for an excellent analysis of these systems.

2. A Research Agenda

Although research using Latent Semantic Analysis (LSA) to assess essays automatically shows promising results [4, 7, 8, 11, 14, 17-19], not enough research has been done on using LSA for instructional software [19]. Previous studies involved both young students and university-age students, and several different knowledge domains; none of them involved the domain of computer science. An open question is how LSA can be used to improve the learning of university-age, computer science students. The kinds of corpora produced by novice programmers have entirely different characteristics than do the customary expository writing texts usually studied by LSA researchers.

The research aims to combine LSA's ability to provide assessment with the computational power of the Grid, thus avoiding the huge computational demands mentioned by many as a drawback of the method. The goal of the research is to demonstrate that a system based on Latent Semantic Analysis can determine the gaps and/or misconceptions in a computer science student's mental model of the subject thereby allowing immediate and specific content-based feedback customized to a student's personal needs. The initial pilot study, described in section 3, shows promising results.

2.1. LSA and the Grid

The nature of the Grid offers several applications for LSA-related computer science research; exploiting its potentially huge computational power for LSA processing is probably the most obvious. Another use results from the fact that students learning to program, regardless of their native languages, produce programs in the same language – a subset of English-like words. International researchers can share and compare program segments from multi-cultural and multi-lingual students without the need for translation. The Grid could be used to store and retrieve the large training corpora needed for LSA. Its characteristic of easy access could allow researchers to keep their corpora up-to-date with the rapidly changing computer science field.

2.2. Some Research Questions

An examination of the existing LSA-related literature reveals certain unresolved issues. Further research will attempt to answer these questions:

- On what corpus should the LSA system for computer science be trained?
- What is a good size for the corpus? This question is particularly pressing as [21] claim that obtaining a large corpus is "the most relevant problem" when automatically assessing essays.
- How do students feel about using a computer-based grading system? Is there any correlation with learning style or cognitive style?
- How can the Grid be used to solve the heavy computational demands of LSA?

3. A Pilot Study

Thomas et al [22] describe a pilot study carried out to test the feasibility of some of the research proposed in this paper. The study involved answers to three essay questions written by computer science students in a graduate course on computer architectures. The study used the Spearman's rho statistical test [23] to compare the marks given by humans to LSA-generated marks. Only one of the three questions showed a statistical correlation between LSA and human marks.

These results are unacceptable for a real-world application because two out of three questions showed insufficient correlation. However, they are encouraging given the extremely small corpus size of only 17 documents, or about 2,000 words for two of the questions and about 600 words for the third question. This pilot study solidified our understanding of how to use LSA, the importance of a large corpus, and how to approach further research to improve the results and increase the applicability of the results of this pilot study.

4. Summary

This paper introduced and briefly explained LSA and stresses that it can be used to provide e-assessment by both formative and summative assessment. It provided

examples of the breadth of existing research that uses LSA for e-assessment. It posed several research questions for which the Grid can be crucial in providing answers. Finally, it discussed a small pilot study conducted to establish the feasibility of the research proposed in this paper.

Acknowledgements

References

[1]   S. A. Cerri, "Human and artificial agent's conversations on the Grid," in *Proceedings of the 1st International LeGE-WG Workshop*, 2002.

[2]   V. Reklaitis, "Shaping e-learning applications for a service oriented Grid," in *Proceedings of the 2nd International LeGE-WG Workshop*, 2003.

[3]   I. Dahn, "Software interoperability problems and e-learning," in *Proceedings of the 2nd International LeGE-WG Workshop*, 2003.

[4]   T. K. Landauer, P. W. Foltz, and D. Laham, "An introduction to Latent Semantic Analysis," *Discourse Processes*, vol. 25, pp. 259-284, 1998.

[5]   S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by Latent Semantic Analysis," *Journal of the American Society for Information Science*, vol. 41, pp. 391-407, 1990.

[6]   T. K. Landauer and S. T. Dumais, "A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition , induction and representation of knowledge," *Psychological Review*, vol. 104, pp. 211-240, 1997.

[7]   D. Wade-Stein and E. Kintsch, "Summary Street: Interactive computer support for writing," University of Colorado, USA 2003.

[8]   T. Miller, "Essay assessment with Latent Semantic Analysis," *Journal of Educational Computing Research*, vol. 28, 2003.

[9]   G. W. Furnas, L. M. Gomez, T. K. Landauer, and S. T. Dumais, "Statistical semantics: How can a computer use what people name things to guess what things people mean when they name things?," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*: ACM, 1982, pp. 251-253.

[10]   G. W. Furnas, S. Deerwester, S. T. Dumais, T. K. Landauer, R. A. Harshman, L. A. Streeter, and K. E. Lochbaum, "Information retrieval using a singular value decomposition model of latent semantic structure," *ACM*, pp. 465-480, 1988.

[11]   P. W. Foltz, "Latent semantic analysis for text-based research," *Behavior Research Methods, Instruments and Computers*, vol. 28, pp. 197-202, 1996.

[12]   R. Soto, "Learning and performing by exploration: Label quality measured by Latent Semantic Analysis," in *Proceedings of CHI '99*. Pittsburgh, PA: ACM, 1999, pp. 418-425.

[13]   C. Wu, M. Berry, S. Shivakumar, and J. McLarty, "Neural networks for full-scale protein sequence classification: Sequence encoding with singular value decomposition," *Machine Learning*, vol. 21, pp. 177-193, 1995.

[14]   G. Chung and G. O'Neil, "Methodological approaches to online scoring of essays," Los Angeles 1997.

[15]   J. R. Skoyles, "Autistic language abnormality: Is it a second-order context learning defect? - The view from Latent Semantic Analysis," in *Proceedings Child Language Seminar*, I. Barriere, Chiat, S. G. Morgan, and B. Woll, Eds. London, 1999, pp. 1.

[16]   R. S. Campbell and J. W. Pennebaker, "The secret life of pronouns: Flexibility in writing style and physical health," *Psychological Science*, vol. 14, pp. 60-65, 2003.

[17]   P. W. Foltz, D. Laham, and T. K. Landauer, "Automated Essay Scoring: Applications to Educational Technology," in *Proceedings of EdMedia '99*, 1999.

[18]   D. J. Steinhart, "Summary Street: An intelligent tutoring system for improving student writing through the use of Latent Semantic Analysis," Department of Psychology, University of Colorado, Boulder, PhD Dissertation 2001.

[19]   B. Lemaire and P. Dessus, "A system to assess the semantic content of student essays," *Journal of Educational Computing Research*, vol. 24, pp. 305-320, 2001.

[20]   P. Wiemer-Hastings, K. Wiemer-Hastings, and A. C. Graesser, "Improving an intelligent tutor's comprehension of students with Latent Semantic Analysis," in *Artificial Intelligence in Education*, S. P. Lajoie and M. Vivet, Eds. Amsterdam: IOS Press, 1999.

[21]   S. Valenti, F. Neri, and A. Cucchiarelli, "An Overview of Current Research on Automated Essay Grading," *Journal of Information Technology Education*, vol. 2, 2003.

[22]   P. Thomas, D. Haley, A. De Roeck, and M. Petre, "E-Assessment using Latent Semantic Analysis in the Computer Science Domain: A Pilot Study," in *Proceedings of the eLearning for Computational Linguistics and Computational Linguistics for eLearning Workshop at COLING 2004.* Geneva, 2004.

[23]   D. C. Howell, *Statistical Methods for Psychology*, Fourth ed. Belmont, CA: Duxbury Press, 1997.