

Latent Semantic Analysis: Theory, Use and Applications

Simon Dennis (Simon.Dennis@colorado.edu)
Tom Landauer (Tom.Landauer@colorado.edu)
Walter Kintsch (Walter.Kintsch@colorado.edu)
Jose Quesada (Jose.Quesada@colorado.edu)
Institute of Cognitive Science, University of Colorado
Boulder, Co 80301 USA

What is LSA?

Latent Semantic Analysis (LSA) 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 (Landauer and Dumais, 1997). The underlying idea is that the aggregate of all the word contexts in which a given word does and does not appear provides a set of mutual constraints that largely determines the similarity of meaning of words and sets of words to each other. These constraints can be solved using linear algebra methods, in particular, the Singular Value Decomposition.

The adequacy of LSA's reflection of human knowledge has been established in a variety of ways. For example, its scores overlap those of humans on standard vocabulary and subject matter tests; it mimics human word sorting and category judgments; it simulates word-word and passage-word lexical priming data; and it accurately estimates passage coherence, learnability of passages by individual students, and the quality and quantity of knowledge contained in an essay. In addition, LSA has found application in a number of areas including selecting educational materials for individual students, guiding online discussion groups, diagnosing mental disorders from prose, matching jobs with candidates and facilitating automated tutors.

About the tutorial

Session 1: In this session, we introduce LSA and cover the basic mathematical foundations.

Session 2: The LSA group maintains a website that allows researchers to use the technique in their work. In this session, we acquaint attendees with the use of the site and ensure that they appreciate the significance of parameter choices they may make. Upon completion attendees will be able to use the site to conduct their own LSA-related research.

Session 3: LSA is both a cognitive theory and an information technology. In this session, we first overview LSA as a cognitive model. Then we discuss some of the many applications, including scoring student essays, information retrieval, cross language retrieval, educational material selection, comprehensibility evaluation and psychiatric assessment.

Bibliographic Sketches

Simon Dennis is a computer scientist and psychologist and currently holds a research professorship at the University of Colorado, Boulder. His prior work has focused on empirical and computational modeling approaches to episodic memory. More recently he has been investigating statistical models of sentence comprehension.

Tom Landauer is Professor of Psychology and Fellow of the Institute of Cognitive Science at the University of Colorado, Boulder. He is also president of Knowledge Analysis Technologies, a company that supplies web-based applications of LSA in education, personnel, and tools for information workers. As Director of Cognitive Science Research at Bellcore in the 1990s he was closely involved in the original research and development of LSA, and is a named inventor on several patents for its application. His research over the last decade has focused on interlocking improvements in LSA and its use as cognitive theory, cognitive science research tool, and practical information technology.

Walter Kintsch is Professor of Psychology and Director of the Institute of Cognitive Science at the University of Colorado in Boulder. His research focus has been on the study of how people understand language, using both experimental methods and computational modeling techniques. In cooperation with the Dutch linguist Teun van Dijk, he formulated the first psychological process theory of discourse comprehension in 1978. In 1988, this work was reformulated as a constraint-satisfaction process. His latest book "Comprehension" appeared in 1998 and argues that many cognitive processes can be usefully conceptualized as comprehension processes. Kintsch is a member of the National Academy of Education and received the Distinguished Scientific Contribution Award of the American Psychological Association in 1992 and an honorary doctorate from the Humboldt University in Berlin in 2001.

Jose Quesada recently received his PhD in Cognitive Science from the University of Grenada. His thesis work focused on the extending the use of LSA principles to problem solving domains including simulated fire fighting and process control as well as aircraft landing efficiency.