Using High-dimensional Semantic Spaces Derived from Large Text Corpora

(A Symposium at the 1995 Annual Cognitive Science Meeting)

Curt Burgess
Department of Psychology
1419 Life Sciences Building
University of California, Riverside
Riverside, CA 92521-0426
curt@cassandra.ucr.edu

Gary Cottrell
Institute for Neural Computation
Computer Science and Engineering Dept. 0114
University of California, San Diego
La Jolla, CA 92093-0114
gary@cs.ucsd.edu

Abstract

Attempting to derive models of semantic memory using psychometric techniques has a long history in cognitive psychology dating back at least to Osgood (1957). Many others have used multidimensional scaling on human judgements of similarity (e.g., Shepard, 1962, 1974; Rips, Shoben, & Smith, 1973; Schvaneveldt, 1990). Recently, a small group of investigators have been using large corpora, 1 million to 500 million words, to develop cognitively plausible high-dimensional semantic models without the need for human judgements on stimuli. These models have become increasingly better at explaining a wide range of cognitive phenomena as they move beyond simple co-occurrence statistics.

Introduction

Forty years ago at the University of Colorado (May 12-14, 1955), a symposium was convened to discuss "Contemporary Approaches to Cognition" (Bruner, 1957). A smaller more specialized meeting such as this was crucial, according to Karl Muenzinger (1957), since the annual meeting of the APA had grown too large. Issues under debate included the tenability of structural vs associationist accounts of cognition. It was at this meeting that Charles Osgood (1957) presented the beginnings of his theory of high-dimensional semantic space which was derived from his semantic differential measuring technique. He notes that this was a line of research that simply would not have been possible without the latest in high-speed computing (ILLIAC, the Illinois digital computer).

Forty years later at the University of Pittsburgh (July 22-25, 1995), a symposium was convened to discuss contemporary approaches to cognition. A smaller more specialized meeting such as this was crucial, according to Curt Burgess and Gary Cottrell, since the annual meeting of the -fill in your professional meeting name here-- had grown too large. Issues under debate included the tenability of structural vs. associationist accounts of cognition. It was at this meeting that a small group of investigators presented the beginnings of their approaches to developing high-dimensional semantic space which was derived from methods from computational linguistics and cognitive science. It was noted that this was a line of research that simply would not have been possible with the latest in high-speed computing.

How times have changed...

High-dimensional semantic space models are beginning to account for a wide range of phenomena: synonym judgements, semantic constraints in parsing, various semantic priming effects, lexical ambiguity retrieval, the typicality effect, lexical/semantic access, as well as other memory and language-based data. These high-dimensional semantic models have implications for theories of concept and language acquisition, as well as information storage. Furthermore, there is a close linkage between applied and theoretical aspects of knowledge representation, particularly as it involves information retrieval and storage systems. One advantage of the corpus-based research discussed in this symposium is the ability to rely on computationally intensive procedures to build large-scale semantic representational systems. These models acquire conceptual representations without supervision with the resulting semantic systems becoming increasingly plausible from a cognitive perspective. Over the last 30 years, the range of phenomena to be dealt with has expanded, the methodological tools have become more sophisticated, the representational structures are richer, and there has been a sharp rise in interdisciplinary efforts.

Times have changed...

At this symposium, we ask those in attendance to do what Charles Osgood once requested ... "imagine a hypothetical semantic space of some unknown number of dimensions" (Osgood, 1971, pg. 7).

Symposium Participants

The following papers were presented in this symposium:

The Promise of High-dimensional Semantics
Curt Burgess, University of California, Riverside

Constructing High-dimensional Semantic Spaces: Mysteries Revealed
Kevin Lund, University of California, Riverside

Hyperspace Analogue to Language (HAL): A General Model of Semantic Memory
Curt Burgess, University of California, Riverside

Words and their Contextualizations Corpus-based Representations of Context
Hinrich Schuetze, Center for the Study of Language and Information, Stanford University

Latent Semantic Analysis as a Model of Memory and Knowledge Representation
Thomas Landauer, University of Colorado

Lexical Access with Internet Semantics
Dan Clouse & Gary Cottrell, University of California, San Diego

High-dimensional Semantics: Lost in vector space?
Gary Cottrell, University of California, San Diego

References

Bruner, J. (Ed.). (1957). *Contemporary approaches to cognition*. Cambridge, MA: Harvard University Press.

Muenzinger, K. Introduction (pp. 1-4). In J. Bruner (Ed.), *Contemporary approaches to cognition.* Cambridge, MA: Harvard University Press.

Osgood, C.E. (1957). A behavioristic analysis of perception and language as cognitive phenomena (pp. 75-118). In J. Bruner (Ed.), *Contemporary approaches to cognition*. Cambridge, MA: Harvard University Press.

Osgood, C.E. (1971). Exploration in semantic space: A personal diary. *Journal of Social Issues, 27*, 5-64.

Shepard, R.N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika, 27*, 125-140.

Shepard, R.N. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrika, 39*, 373-421.

Rips, L.J., Shoben, E.J., & Smith, E.E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior, 12*, 1-20.

Schvaneveldt, R.W. (1990). *Pathfinder associative networks: Studies in knowledge organization*. Norwood, NJ: Ablex.

Smith, E.E., Shoben, E.J., & Rips, L.J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review, 81*, 214-241.