Acquiring and representing distributional information

Instructor:

Magnus Sahlgren

Abstract

There is a rich and growing literature of work in corpus-based and computational linguistics based on the distributional hypothesis in language. A large body of work in language technology uses distributional information to compute semantic similarities between words. Various techniques are employed to translate distributional data into semantic representations and to clarify what kind of semantic knowledge is acquired through distributional evidence. This course will cover the theoretical basis of the distributional methodology, and will give an overview of the basics of methods, techniques and measures used for distributional semantic modelling. The course will also exemplify how distributional analysis can be used in quantitative research, and students will have the opportunity to try distributional methods hands-on in lab assignments.

Literature:

Landauer, T. and S. Dumais (1997) "A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge", Psychological Review, 104 (2), pp. 211–240 (http://lsa.colorado.edu/papers/plato/plato.annote.html)

Lund, K., C. Burgess and R. Atchley (1995) "Semantic and associative priming in high-dimensional semantic space", Proceedings of the 17th Annual Conference of the Cognitive Science Society, pp. 660–665 (locutus.ucr.edu/reprintPDFs/lba95csp.pdf)

Sahlgren, M. (2006) "The Distributional Hypothesis", Rivista di Linguistica (Italian Journal of Linguistics), 20 (1). pp. 33-53 (http://soda.swedish-ict.se/3941/1/sahlgren.distr-hypo.pdf) Widdows, D. (2004) "Geometry and Meaning", CSLI Publications. (Chapter 5, Word Vectors and Search Engines: http://www.puttypeg.net/papers/vector-chapter.pdf)

Schütze, H. (1992) "Dimensions of Meaning", Supercomputing '92 pp. 787-796 (www2.parc.com/istl/groups/qca/papers/Schuetze-Meaning92.ps.gz)

Sahlgren, M. (2005) "An Introduction to Random Indexing", Methods and Applications of Semantic Indexing Workshop at the 7th International Conference on Terminology and Knowledge Engineering, 16 Aug 2005, Copenhagen, Denmark (http://soda.swedish-ict.se/221/1/RI_intro.pdf)

Turney, P. D. and P. Pantel (2010) "From Frequency to Meaning: Vector Space Models of Semantics", Journal of Artificial Intelligence Research, Volume 37, pp. 141-188 (http://www.jair.org/media/2934/live-2934-4846-jair.pdf)