A Guided Tour of Latent Semantic Mapping

Originally formulated in the context of information retrieval, latent semantic analysis exhibits three main characteristics: (i) words and documents (i.e., discrete entities) are mapped onto a continuous vector space; (ii) this mapping is determined by global correlation patterns; and (iii) dimensionality reduction is an integral part of the process. Because such fairly generic properties may be advantageous in a variety of different contexts, this has sparked interest in a more inclusive interpretation of the underlying paradigm. The outcome is latent semantic mapping, a data-driven framework for modeling global relationships implicit in large volumes of data. The purpose of this talk is to give a broad overview of the framework, highlight the possibilities it offers for general feature extraction, and underscore the multi-faceted benefits it can bring to a number of problems in speech and language processing. We conclude with a discussion of the inherent trade-offs associated with the approach, and some perspectives on its likely role in information extraction going forward.

Jerome R. Bellegarda is currently Apple Distinguished Scientist in Human Language Technologies at Apple Inc, Cupertino, California. His general interests span voice-driven man-machine communications, multiple input/output modalities, and multimedia knowledge management. In these areas he has written approximately 150 publications, and holds about 50 U.S. and foreign patents. He has served on many international scientific committees, review panels, and advisory boards. In particular, he has worked as Expert Advisor on speech technology for both the National Science Foundation and the European Commission, was Associate Editor for the IEEE Transactions on Audio, Speech and Language Processing, served on the IEEE Signal Processing Society Speech Technical Committee, and is currently an Editorial Board member for both Speech Communication and the ACM Transactions on Speech and Language Processing. He is a Fellow of the IEEE.

*Find more information about the talk and LTI colloquium at: http://www.cs.cmu.edu/afs/cs.cmu.edu/project/cmt-55/lti/Courses/700/*