Latent Semantic Mapping: Principles & Applications

Latent Semantic Mapping: Principles & Applications

Jerome R. Bellegarda

ISBN: 978-3-031-01428-4 paperback
ISBN: 978-3-031-01428-4 paperback

ISBN: 978-3-031-02556-3 ebook
ISBN: 978-3-031-02556-3 ebook

DOI: 10.1007/978-3-031-02556-3

A Publication in the Springer series

SYNTHESIS LECTURES ON SPEECH AND AUDIO PROCESSING #3

Lecture #3

Series Editor: B.H. Juang, Georgia Tech

Library of Congress Cataloging-in-Publication Data

Series ISSN: 1932-121X print
Series ISSN: 1932-1678 electronic

First Edition

10 9 8 7 6 5 4 3 2 1

Latent Semantic Mapping: Principles & Applications

Jerome R. Bellegarda
Apple Inc.

SYNTHESIS LECTURES ON SPEECH AND AUDIO PROCESSING #3

ABSTRACT

Latent semantic mapping (LSM) is a generalization of latent semantic analysis (LSA), a paradigm originally developed to capture hidden word patterns in a text document corpus.

In information retrieval, LSA enables retrieval on the basis of conceptual content, instead of merely matching words between queries and documents. It operates under the assumption that there is some latent semantic structure in the data, which is partially obscured by the randomness of word choice with respect to retrieval. Algebraic and/or statistical techniques are brought to bear to estimate this structure and get rid of the obscuring "noise." This results in a parsimonious continuous parameter description of words and documents, which then replaces the original parameterization in indexing and retrieval.

This approach exhibits three main characteristics:

- *discrete* entities (words and documents) are mapped onto a *continuous* vector space;
- this mapping is determined by *global correlation patterns*; and
- *dimensionality reduction* is an integral part of the process.

Such fairly generic properties are advantageous in a variety of different contexts, which motivates a broader interpretation of the underlying paradigm. The outcome (LSM) is a data-driven framework for modeling meaningful global relationships implicit in large volumes of (not necessarily textual) data.

This monograph gives a general overview of the framework, and underscores the multi-faceted benefits it can bring to a number of problems in natural language understanding and spoken language processing. It concludes with a discussion of the inherent tradeoffs associated with the approach, and some perspectives on its general applicability to data-driven information extraction.

KEYWORDS

natural language processing, long-span dependencies, data-driven modeling, parsimonious representation, singular value decomposition.

Contents

List of Figures

List of Tables