Abstract

Latent semantic analysis (LSA) is a technique that analyzes relationships between documents and its terms, and it discovers a data representation that has a lower dimension than the original semantic space. Essentially, the reduced dimensionality preserves the most crucial aspects of the data since LSA analyzes documents to find latent meaning in the corpus. The latent semantic space is determined by singular value decomposition (SVD), which enables a powerful process to simplify any rectangular matrix into a product of three unique components. The purpose of using SVD is to retrieve a sufficient amount of dimensions, which reveal a relevant structure that spans the original term-document matrix. In this study, LSA was used to find particular associations with user queries in a sample of documents from Medline Industries, Inc. Selecting an appropriate dimension for a reduced representation is suitable to represent the original latent space. The reduced model of the term-document matrix shows that SVD is capable of dealing with semantic problems. Overall, the goal is to overcome the problem of unsatisfactory indexed results by revealing hidden relationships among the terms and documents.

Index Terms

Computer Science Information Science

Keywords
Latent Semantic Analysis  Singular Value Decomposition  Text Mining  Dimensionality Reduction