Latent Semantic Indexing

Seminar “Theoretical Topics in Data Science”

Vahe Eminyan

vahe.eminyan@rwth-aachen.de

15.11.2023

Overview

Introduction

LSI Background

Original Paper Overview and Emphasized Aspect

LSI by Random Projection

References

In this section

- Motivation
- What is LSI (rough idea)

In this section

- SVD explanation
- How does LSI work

Original Paper Overview and Emphasized Aspect

In this section

- Two interesting questions Papadimitriou et al. investigated [1]
  - Why does LSI perform well (why does it find the documents semantically related to each other)
  - How can we speed up the computation
- We will focus on the second question

In this section

- In this section we will investigate the question ”How we can speed up the computation”: Informal formulation of the main theorem of this section (Theorem 5 original paper)
- Introduction of theorems and lemmas that are necessary for the proof of the main theorem
- Introduction: the main theorem (Theorem 5 original paper)
- Proof of the main theorem (Theorem 5 original paper)
- Computational savings achieved by LSI by random projection

References

Christos H. Papadimitriou, Prabhakar Raghavan, Hisao Tamaki, and Santosh Vempala.
Latent semantic indexing: A probabilistic analysis.
Journal of Computer and System Sciences, 61(2):217–235, 2000.
URL: https://www.sciencedirect.com/science/article/pii/S0022000000917112, doi:10.1006/jcss.2000.1711.

References

The End