

ALTERNATE LOW RANK APPROXIMATION IN LATENT SEMANTIC ANALYSIS

HASAN ERBAY, FATİH VARÇIN and FAHRETİN HORASAN

Computer Engineering Department, Kırıkkale University

Yahşihan, 71450 Kırıkkale, Turkey

E-mail: hxe68@yahoo.com

Latent semantic analysis (LSA) is a way of discovering hidden concepts between terms and documents or within the document collection. Each document and term is expressed as a vector with elements corresponding to these concepts to form a term-document matrix. LSA uses a low rank approximation to the term-document matrix in order to remove irrelevant information and extract more important relations. The irrelevant information is called as noise and does not have a noteworthy effect on the meaning of the document collection. The singular value decomposition (SVD) has been the main tool obtaining the low rank approximation in LSA. Due to the increase in the document collection, we need to renew the approximation. This can be done via folding-in, re-computing SVD or updating the SVD. However, the computational time of re-computing or updating the SVD of the term-document matrix is high when adding new terms and/or documents to pre-existing document collection. Therefore, this problem opened the door of using other matrix decompositions for LSA as ULV and URV decompositions. This study shows that the truncated ULV decomposition is a good alternative to the SVD in LSA modelling.

REFERENCES

- [1] S. Deerwester, S. Dumais, S. Furnas, T. G. Landauer and R. Harshman. Indexing by latent semantic analysis. *Journal of the American society for information science*, **41** (2):126-138, 1990.
- [2] M. W. Berry and R. D. Fierro . Low-Rank Orthogonal Decompositions for Information Retrieval Applications . *Numerical Linear Algebra with Applications* , **3** 301-328, 1996.
- [3] M. W. Berry, Z. Drmac and E. Jessup. Matrices, vector spaces and information retrieval. *SIAM Review*, **41** (2):335-362, 1999.
- [4] J. Barlow and H. Erbay. Modifiable Low-Rank Approximation to a Matrix. *Numerical Linear Algebra with Applications*, **16** (10):833-860, 2009.
- [5] K. Lochbaum ve L. Streeter. Comparing and combing the effectiveness of latent semantic indexing and the ordinary vector space model for information retrieval. *Information Processing and Management*, **55** (6):665-675, 1989.
- [6] S. Dumais. Improving the retrieval of information from external sources. *Behavior Research Methods, Instrument and Computers*, **23** (2):229-236, 1991.
- [7] E. Jessup and J. Martin. Taking a new look at the latent semantic analysis approach to information retrieval. *Computational information retrieval*, 121-144, 2001.
- [8] Å. Björck. *SIAM publications*. Philadelphia, PA, 1996.