Low-rank matrix approximation with weights or missing data is NP-hard

Nicolas Gillis and François Glineur

Université catholique de Louvain, Belgium

Abstract

Weighted low-rank approximation (WLRA) [1], a dimensionality reduction technique for data analysis, has been successfully used in several applications, such as in collaborative filtering to design recommender systems [3] or in computer vision to recover structure from motion [4]. WLRA can be formulated as follows: given an m-by-n real matrix M, an m-by-n nonnegative real matrix W and a factorization rank r, find an m-by-r real matrix U and an n-by-r real matrix V such that $||M - UV^T||_W^2 = \sum_{ij} W_{ij}(M - UV^T)_{ij}^2$ is minimized.

In this talk, we show that computing an optimal weighted low-rank approximation is NP-hard, already when a rank-one approximation is sought (i.e., for $r = 1$). In fact, we show that it is hard to compute approximate solutions to the WLRA problem with some prescribed accuracy [2]. Our proofs are based on reductions from the maximum-edge biclique problem, and apply to strictly positive weights as well as to binary weights (the latter corresponding to low-rank matrix approximation with missing data).

Keywords

Weighted low-rank approximation, Missing data, Matrix completion with noise, PCA with missing data, Computational complexity, Maximum-edge biclique problem.

References

[1] Gabriel, K.R. and S. Zamir (1979). Lower rank approximation of matrices by least squares with any choice of weights. *Technometrics 21(4)*, 489–498.

[2] Gillis, N. and F. Glineur (2010). Low-rank matrix approximation with weights or missing data is NP-hard. *arXiv:1012.0197v1*.

[3] Koren, Y., R. Bell, and C. Volinsky (2009). Matrix factorization techniques for recommender systems. *IEEE Computer 42(8)*, 30–37.

[4] Shum, H., K. Ikeuchi, and R. Reddy (1995). Principal component analysis with missing data and its application to polyhedral object modeling. *IEEE Trans. Pattern Anal. Mach. Intelligence 17(9)*, 854–867.