Adachi, Kohei; Trendafilov, Nickolay T.
Sparsest factor analysis for clustering variables: a matrix decomposition approach. (English)
Zbl 1416.62319
Adv. Data Anal. Classif., ADAC 12, No. 3, 559-585 (2018).

Summary: We propose a new procedure for sparse factor analysis (FA) such that each variable loads only one common factor. Thus, the loading matrix has a single nonzero element in each row and zeros elsewhere. Such a loading matrix is the sparsest possible for certain number of variables and common factors. For this reason, the proposed method is named sparsest FA (SSFA). It may also be called FA-based variable clustering, since the variables loading the same common factor can be classified into a cluster. In SSFA, all model parts of FA (common factors, their correlations, loadings, unique factors, and unique variances) are treated as fixed unknown parameter matrices and their least squares function is minimized through specific data matrix decomposition. A useful feature of the algorithm is that the matrix of common factor scores is re-parameterized using QR decomposition in order to efficiently estimate factor correlations. A simulation study shows that the proposed procedure can exactly identify the true sparsest models. Real data examples demonstrate the usefulness of the variable clustering performed by SSFA.

MSC:

62H25 Factor analysis and principal components; correspondence analysis
62H30 Classification and discrimination; cluster analysis (statistical aspects)
15A23 Factorization of matrices

Keywords:

exploratory factor analysis; sparsest loadings; matrix decomposition factor analysis; variable clustering; QR re-parameterization

Software:

sparsenet

Full Text: DOI Link

References:

[1] Adachi, K., Some contributions to data-Fitting factor analysis with empirical comparisons to covariance-Fitting factor analysis, J Jpn Soc Comput Stat, 25, 25-38, (2012) · Zbl 1326.62134 · doi:10.5183/jjscs.1106001$_1$97

[2] Adachi, K., A matrix-intensive approach to factor analysis, Jpn J Stat, 44, 363-382, (2014)

[3] Adachi, K.; Trendafilov, NT; Carpita, M. (ed.); Brentari, E. (ed.); Qannari, EM (ed.), Sparse orthogonal factor analysis, 227-239, (2014), Heidelberg

[4] Aggarwal CC (2015) Data mining: the textbook. Springer, New York · Zbl 1311.68001 · doi:10.1007/978-3-319-14142-8

[5] Costa PT, McCrae RR (1992) NEO PI-R professional manual: revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO-FFI). Psychological Assessment Resources, Odessa

[6] Leeuw, J.; Montfort, K. (ed.); Oud, J. (ed.); Satorra, A. (ed.), Least squares optimal scaling of partially observed linear systems, 121-134, (2004), Dordrecht · Zbl 05196650 · doi:10.1007/978-1-4020-1958-6$_7$

[7] Eldén L (2007) Matrix methods in data mining and pattern recognition. SIAM, Philadelphia · Zbl 1120.68092 · doi:10.1137/1.9780898718867

[8] Everitt BS (1993) Cluster analysis, 3rd edn. Edward Arnold, London · Zbl 0507.62060

[9] Gan G, Ma C, Wu J (2007) Data clustering: theory, algorithms, and Applications. Society of Industrial and Applied Mathematics (SIAM), Philadelphia · Zbl 1185.68274 · doi:10.1137/1.9780898718348

[10] Goldberg, LR, The development of markers for the big-five factor structure, Psychol Assess, 4, 26-42, (1992)· doi:10.1037/1040-3590.4.1.26

[11] Harman HH (1976) Modern factor analysis, 3rd edn. The University of Chicago Press, Chicago · Zbl 0095.13403

[12] Hirose, K.; Yamamoto, M., Estimation of an oblique structure via penalized likelihood factor analysis, Comput Stat Data Anal, 79, 120-132, (2014) · Zbl 06984059 · doi:10.1016/j.csda.2014.05.011

[13] Hirose K, Yamamoto M (2014b) Sparse estimation via nonconcave penalized likelihood in factor analysis model. Comput,

Statist. doi:10.1007/s11222-014-9475-z · Zbl 1332.62194 · doi:10.1007/s11222-014-9475-z

[14] Holzinger KJ, Swineford F (1939) A study in factor analysis: the stability of a bi-factor solution. University of Chicago, Supplementary Educational Monographs, No. 48

[15] Izenman AJ (2008) Modern multivariate statistical techniques: regression, classification, and manifold learning. Springer, New York · Zbl 1155.62040 · doi:10.1007/978-0-387-78189-1

[16] Jolliffe, IT; Trendafilov, NT; Uddin, M., A modified principal component technique based on the LASSO, J Comput Graph Stat, 12, 531-547, (2003)· doi:10.1198/1061860032148

[17] Knowles, D.; Ghahramani, Z., Nonparametric Bayesian sparse factor models with applications to gene expression modeling, Ann Appl Stat, 5, 1534-1552, (2011) · Zbl 1223.62013 · doi:10.1214/10-AOAS435

[18] Mazumder, R.; Friedman, J.; Hastie, T., Sparsenet: coordinate descent with nonconvex penalties, J Am Stat Assoc, 106, 1125-1138, (2011) · Zbl 1229.62091 · doi:10.1198/jasa.2011.tm09738

[19] Mulaik SA (2010) Foundations of factor analysis, 2nd edn. CRC Press, Boca Raton · Zbl 1188.62185

[20] Rattray, M.; Stegle, O.; Sharp, K.; Winn, J., Inference algorithms and learning theory for Bayesian sparse factor analysis, J Phys Conf Ser, 197, 1-10, (2009)· doi:10.1088/1742-6596/197/1/012002

[21] Reyment R, Jöreskog KG (1996) Applied factor analysis in the natural sciences. Cambridge University Press, Cambridge · Zbl 0868.62051

[22] Sampson RJ (1968) R-mode factor analysis program in FORTRAN II for the IBM 1620 computer. Kansas Geol Survey Comput Contrib 20

[23] Seber GAF (2008) A matrix handbook for statisticians. Wiley, Hoboken · Zbl 1143.15001

[24] Sočan G (2003) The incremental value of minimum rank factor analysis. Ph.D. Thesis, University of Groningen, Groningen

[25] Spearman, C., 'general intelligence' objectively determined and measured, Am J Psychol, 15, 201-293, (1904)· doi:10.2307/1412107

[26] Stegeman, A., A new method for simultaneous estimation of the factor model parameters, factor scores, and unique parts, Comput Stat Data Anal, 99, 189-203, (2016) · Zbl 1468.62181 · doi:10.1016/j.csda.2016.01.012

[27] Berge, JMF, A generalization of kristof's theorem on the trace of certain matrix products, Psychometrika, 48, 519-523, (1983) · Zbl 0536.62093 · doi:10.1007/BF02293876

[28] ten Berge JMF (1993) Least squares optimization in multivariate analysis. DSWO Press, Leiden

[29] Trendafilov, NT, From simple structure to sparse components: a review, Comput Stat, 29, 431-454, (2014) · Zbl 1306.65143 · doi:10.1007/s00180-013-0434-5

[30] Trendafilov, NT; Unkel, S., Exploratory factor analysis of data matrices with more variables than observations, J Comput Graph Stat, 20, 874-891, (2011)· doi:10.1198/jcgs.2011.09211

[31] Trendafilov, Nickolay T.; Unkel, Steffen; Krzanowski, Wojtek, Exploratory factor and principal component analyses: some new aspects, Statistics and Computing, 23, 209-220, (2011) · Zbl 1322.62043 · doi:10.1007/s11222-011-9303-7

[32] Unkel, S.; Trendafilov, NT, Simultaneous parameter estimation in exploratory factor analysis: an expository review, Int Stat Rev, 78, 363-382, (2010)· doi:10.1111/j.1751-5823.2010.00120.x

[33] Vichi, M.; Saporta, G., Clustering and disjoint principal component analysis, Comput Stat Data Anal, 53, 3194-3208, (2009) · Zbl 1453.62230 · doi:10.1016/j.csda.2008.05.028

[34] Zaki MJ, Meira W (2014) Data mining and analysis: fundamental concepts and algorithms. Cambridge University Press, Cambridge · Zbl 1331.68005 · doi:10.1017/CBO9780511810114

[35] Zou, DM; Hastie, T.; Tibshirani, R., Sparse principal component analysis, J Comput Graph Stat, 15, 265-286, (2006)· doi:10.1198/106186006X113430