

Abstract

J. Chemometrics, **24**, 209-217 (2010)

Varmuza K., Filzmoser P., Liebmann B.:

Random projection experiments with chemometric data

Random projection (RP) is a linear method for the projection of high-dimensional data onto a lower dimensional space. RP uses projection vectors (loading vectors) that consist of random numbers taken from a symmetric distribution with zero mean; many successful applications have been reported for high-dimensional data sets.

The basic ideas of RP are presented, and tested with artificial data, data from chemoinformatics and from chemometrics. RP's potential in dimensionality reduction is investigated by a subsequent cluster analysis, classification or calibration, and is compared to PCA as a reference method. RP allowed drastic reduction in data size and computing time, while preserving the performance quality.

Successful applications are shown in structure similarity searches (53 478 chemical structures characterized by 1233 binary substructure descriptors) and in classification of mutagenicity (6506 chemical structures characterized by 1455 molecular descriptors). Only in calibration tasks with low-dimensional data as in many chemical applications, RP showed limited performance.

For special applications in chemometrics with very large data sets and/or severe restrictions for hardware and software resources, RP is a promising method.