Chemometrics and Intelligent Laboratory Systems 143 (2015) 28–39

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Quality prediction based on HOPLS-CP for batch processes

Lijia Luo *, Shiyi Bao, Zengliang Gao

College of Mechanical Engineering, Zhejiang University of Technology, Engineering Research Center of Process Equipment and Remanufacturing, Ministry of Education, Hangzhou, China

ARTICLE INFO

Article history:
Received 20 October 2014
Received in revised form 6 February 2015
Accepted 7 February 2015
Available online 14 February 2015

Keywords:
Tensor
Multilinear regression
Higher-order partial least squares
CP decomposition
Quality prediction
Batch processes

ABSTRACT

A new multilinear regression method, called the higher-order partial least squares based on CP (canonical decomposition/parallel factor analysis) decomposition (HOPLS-CP), is proposed. Unlike the unfold-PLS (i.e., MPLS) which needs to unfold tensors into matrices for modeling, HOPLS-CP maintains the tensorial representation of data by using the tensor decomposition. Therefore, HOPLS-CP avoids the drawbacks induced by the data unfolding operation and can summarize the multi-way interaction in the data. HOPLS-CP ensures a maximize correlation between two tensors by projecting them into a common latent subspace and independent loading subspaces, where the optimal latent vectors and loading vectors are obtained by performing the CP decomposition on the cross-covariance tensor of two tensors. A generalized HOPLS-CP model is then developed to build a regression model for uneven tensors with varying sizes on one mode. An outstanding advantage of GHOPLS-CP is that it naturally solves the uneven data length problem with no need for data alignment, guaranteeing a better modeling ability and the intuitive interpretability. HOPLS-CP and GHOPLS-CP are applied for the quality prediction in a benchmark fed-batch penicillin fermentation process. Their advantages, including well predictive ability and anti-noise capability, are illustrated by two case studies.

1. Introduction

Batch process is playing an important role in producing low-volume and high-value-added products, because of its high flexibility to adapt the rapidly changing market [1,2]. At present, batch processes have been widely used in chemical, biochemical, pharmaceutical and semiconductor industries, and so on [3]. The increasing market competition has aroused the demand of consistent and high quality products for batch processes. However, due to the lack of online measurements of quality variables [4,5], online quality control in batch processes is difficult. This makes the quality prediction becomes a highly necessary part of batch process operations. The main task and objective of quality prediction are to estimate quality variables as fast and accurate as possible, and thus guide the process operation to obtain high quality products.

Quality prediction methods have attracted much attention in last decades. Typically, the data-driven quality prediction method has become a hot topic, owing to the widespread application of process automation techniques in collecting massive process data. Unlike the conventional first-principle model based method [4], the data-driven method directly derives prediction model from process data with little requirement on process mechanism and prior knowledge. This advantage makes data-driven modeling methods be more applicable in real industrial processes. The multi-way partial least squares (MPLS), also termed as unfold-PLS, is one of the most popular data-driven quality prediction methods for batch processes. Lots of extended MPLS methods have been proposed to enhance the prediction performance [6–10]. For example, Duchesne and MacGregor [6] proposed a pathway multi-block PLS method by utilizing intermediate quality measurements to identify time-specific effects of process variables on the final product quality. Lu and Gao [7] developed a phase-based PLS method by taking the multi-phase characteristic of batch processes into account. Yu [8] proposed a multi-way Gaussian mixture model based adaptive kernel partial least-squares (MGMM-AKPLS) method for the quality prediction of nonlinear batch processes.

A common feature of MPLS-based methods is that they need to unfold the three-way batch data array into a matrix to build the process model. However, the data unfolding operation may result in an unfolded dataset with the "large variable number but small sample size" problem. For instance, a three-way array $\underline{\boldsymbol{X}}(10 \times 10 \times 100)$, which has 10 batches, 10 variables and 100 time slices, is unfolded to be a matrix $\boldsymbol{X}(10 \times 1000)$ with 10 samples and 1000 features. If applying PLS on $\boldsymbol{X}(10 \times 1000)$, only ten samples are available to optimize the 1000-dimensional loading vectors, probably resulting in an unreliable estimate of model parameters [11]. Meanwhile, the unfolded data matrix may contain lots of information with high autocorrelation and cross-correlation complexities. These drawbacks bring difficulty to build a superior MPLS model. Besides, MPLS fails to offer an explicit description of the three-way interaction in the data, because it destroys the three-way data structure by data unfolding. Specifically, MPLS does not separately extract loading vectors on variable and time modes of the three-way batch data but merging them together, and thus the interaction between variables and time is not explicit.

* Corresponding author. Tel.: +86 571 88320349.
E-mail address: lijialuo@zjut.edu.cn (L. Luo).

http://dx.doi.org/10.1016/j.chemolab.2015.02.010
0169-7439/

In recent years, the tensor data analysis based on multi-linear algebra has been successfully applied in many fields, such as chemistry, signal processing, neuroscience, graph analysis, computer vision, food industry, and so on [12,13]. Several surveys have summarized recent developments and applications of tensor data analysis methods [12–15]. The strength of tensor data analysis methods is that they maintain the tensorial representation of data, and thus avoid those drawbacks induced by data unfolding. Compared with MPLS, the tensor analysis is able to summarize all information, i.e., all main effects and all interactions together, in the tensor data. Specifically, tensor analysis methods extract the feature on each mode of tensor data through a few components and describe the relations between these components [16].

N-way PLS (N-PLS) [17] and higher-order partial least squares (HOPLS) [11] are two typical multi-linear regression methods. Bro [17] proposed N-PLS as a natural extension of the bilinear PLS [18] to higher orders, combining canonical decomposition/parallel factor analysis (CANDECOMP/PARAFAC or CP) [12,19,20] with PLS. It decomposes independent data and dependent data to a set of score vectors and weight vectors simultaneously, subject to maximum pairwise covariance of score vectors [11,17]. Bro [17] demonstrated some advantages of N-PLS as compared to unfold-PLS, including robustness to noise, stabilized solution, increased predictability and intuitive interpretability. Zhao et al. [11] presented HOPLS based on the block Tucker decomposition. Owing to the better fitness ability of Tucker model over the CP model [12,21], HOPLS has a better predictive ability than N-PLS [11]. However, the multi-fold cross-validation should be used to choose proper tuning parameters for the HOPLS model, which brings a heavy computational burden and is very slow for large-scale data. This can be very annoying in practical applications. Similar to unfold-PLS, neither HOPLS nor N-PLS can be used to modeling the data with uneven sizes, which prevents their further applications.

Most of quality prediction methods for batch processes idealistically assume that all batch data have the same size, ignoring the uneven-duration nature of many batch processes. In fact, the duration of a real batch process is usually not fixed due to unavoidable disturbances and changes of operating conditions, resulting in uneven-length batch data. In such case, different types of trajectory synchronization methods can be adopted to handle the uneven-length problem, such as simply cutting all batch data to the minimum length [22], dynamic time warping (DTW) [23] and correlation optimization warping (COW) [24]. However, the trajectory synchronization operation alters the original data record, which may distort the underlying correlation information of data and reduce the prediction ability. Besides, the physical meaning of time scale may be lost after trajectory synchronization, which reduces the interpretability of models. Therefore, an effective method should be developed to solve the uneven-length problem without aligning batch trajectories.

In this paper, a new multi-linear regression method, called the higher-order partial least squares based on CP decomposition (HOPLS-CP), is proposed to predict the tensor $\underline{\boldsymbol{Y}}$ from a tensor $\underline{\boldsymbol{X}}$. Similar to other tensor analysis methods, HOPLS-CP is able to directly modeling tensor data without using data unfolding. HOPLS-CP aims to simultaneously decompose tensors $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{Y}}$ into a set of vectors, termed as latent vectors and loading vectors, and ensures that the latent vectors from $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{Y}}$ have maximum pairwise covariance. To achieve this goal, HOPLS-CP performs CP decomposition on the cross-covariance tensor of $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{Y}}$ to optimize all loading vectors, and then obtain latent vectors by simple tensor operations. A generalized HOPLS-CP (GHOPLS-CP) method is also developed to build a regression model for uneven tensors with varying sizes on one mode. In the GHOPLS-CP model, the loading vector on the mode with varying sizes is allowed to vary with data sizes, while those loading vectors on other modes are invariable like the HOPLS-CP model. By this way, GHOPLS-CP naturally solves the uneven-length problem with no need for data alignment. HOPLS-CP and GHOPLS-CP are tested in a fed-batch penicillin fermentation process for quality prediction. The results indicate that they have well predictive ability and anti-noise capability.

2. Notation and preliminaries

Tensors (multi-way arrays) are denoted by underlined boldface capital letters, e.g., $\underline{\boldsymbol{X}}$. The order of a tensor is the number of dimensions, namely ways or modes. Matrices are denoted by boldface capital letters, e.g., $\boldsymbol{X}$. Particularly, the identity matrix is denoted by $\boldsymbol{I}$. Vectors are denoted by boldface lowercase letters, e.g., $\boldsymbol{x}$. The ith entry of a vector $\boldsymbol{x}$ is denoted by x_i. The ith column of a matrix $\boldsymbol{X}$ is denoted by $\boldsymbol{x}_i$ and the element (i, j) is denoted by x_{ij}. The element $(i_1, i_2, \ldots, i_N)$ of an Nth-order tensor $\underline{\boldsymbol{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is denoted by $x_{i_1 i_2 \cdots i_N}$. The nth factor vector and matrix in a sequence is denoted by $\boldsymbol{x}^{(n)}$ and $\boldsymbol{X}^{(n)}$, respectively. $\underline{\boldsymbol{X}}_{(n)}$ denotes the mode-n matricization of a tensor $\underline{\boldsymbol{X}}$ [12].

The norm of an Nth-order tensor $\underline{\boldsymbol{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is defined as

$$\|\underline{\boldsymbol{X}}\| = \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} x_{i_1 i_2 \cdots i_N}^2}.$$

The n-mode product of a tensor $\underline{\boldsymbol{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_n \times \cdots \times I_N}$ with a matrix $A \in \mathbb{R}^{J \times I_n}$ is denoted by $\underline{\boldsymbol{Y}} = \underline{\boldsymbol{X}} \times_n A \in \mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times J \times I_{n+1} \times \cdots \times I_N}$ with [12,25]

$$y_{i_1 \cdots i_{n-1} j i_{n+1} \cdots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 \cdots i_n \cdots i_N} a_{j i_n}.$$

The n-mode product of a tensor $\underline{\boldsymbol{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_n \times \cdots \times I_N}$ with a vector $\boldsymbol{t} \in \mathbb{R}^{I_n}$ is denoted by $\underline{\boldsymbol{Y}} = \underline{\boldsymbol{X}} \times_n t \in \mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times I_{n+1} \times \cdots \times I_N}$ with

$$y_{i_1 \cdots i_{n-1} i_{n+1} \cdots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 \cdots i_n \cdots i_N} t_{i_n}.$$

The inner product of two same-sized tensors $\underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}} \in \mathbb{R}^{I_1 \times \cdots \times I_n \times \cdots \times I_N}$ is defined as

$$\langle \underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}} \rangle = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} x_{i_1 i_2 \cdots i_N} y_{i_1 i_2 \cdots i_N}$$

For an Nth-order tensor $\underline{\boldsymbol{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_n \times \cdots \times I_N}$ and an Mth-order tensor $\underline{\boldsymbol{Y}} \in \mathbb{R}^{J_1 \times \cdots \times I_n \times \cdots \times J_M}$ with the same size on the nth-mode, their n-mode cross-covariance is denoted by $\underline{\boldsymbol{Z}} = \langle \underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}} \rangle_{\{n;n\}} \in \mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times I_{n+1} \times \cdots \times I_N \times J_1 \times \cdots \times J_{n-1} \times J_{n+1} \times \cdots \times J_M}$ with

$$z_{i_1 \cdots i_{n-1} i_{n+1} \cdots i_N j_1 \cdots j_{n-1} j_{n+1} \cdots j_M} = \sum_{i_n=1}^{I_n} x_{i_1 \cdots i_n \cdots i_N} y_{j_1 \cdots i_n \cdots j_M}.$$

The Tucker decomposition of an Nth-order tensor $\underline{\boldsymbol{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_n \times \cdots \times I_N}$ can be concisely expressed by the Tucker operator as [12,25]

$$\underline{\boldsymbol{X}} \approx \underline{\boldsymbol{G}} \times_1 \boldsymbol{A}^{(1)} \times_2 \cdots \times_N \boldsymbol{A}^{(N)} \equiv [\![\underline{\boldsymbol{G}}; \boldsymbol{A}^{(1)}, \ldots, \boldsymbol{A}^{(N)}]\!]$$

where $\underline{\boldsymbol{G}} \in \mathbb{R}^{J_1 \times \cdots \times J_n \times \cdots \times J_N}$ is called the core tensor, and $\boldsymbol{A}^{(n)} \in \mathbb{R}^{J_n \times I_n}$ are factor matrices. Likewise, the CP decomposition of an Nth-order tensor $\underline{\boldsymbol{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_n \times \cdots \times I_N}$ can be concisely expressed by the Kruskal operator as [12,25,26]

$$\underline{\boldsymbol{X}} \approx \sum_{r=1}^{R} \boldsymbol{a}_r^{(1)} \circ \boldsymbol{a}_r^{(2)} \circ \cdots \circ \boldsymbol{a}_r^{(N)} \equiv [\![\boldsymbol{A}^{(1)}, \ldots, \boldsymbol{A}^{(N)}]\!]$$

where the symbol "$\circ$" denotes the vector outer product, and $\boldsymbol{a}_r^{(n)} \in \mathbb{R}^{I_n}$ is the rth-column vector of matrix $\boldsymbol{A}^{(N)}$. The Kruskal operator is actually a special case of the Tucker operator where the core tensor is an unit tensor [12,25], i.e.,

$$[\![\boldsymbol{A}^{(1)}, \ldots, \boldsymbol{A}^{(N)}]\!] = [\![\underline{\boldsymbol{J}}; \boldsymbol{A}^{(1)}, \ldots, \boldsymbol{A}^{(N)}]\!]$$