Novel common and special feature extraction method for modeling multi-grade processes

Jingxiang Liu*, Tao Liu*, Junghui Chen **

*Institute of Advanced Control Technology, Dalian University of Technology, Dalian, 116024, P. R. China (e-mail: liujx0505@mail.dlut.edu.cn; liurouter@ieee.org)
** Department of Chemical Engineering, Chung-Yuan Christian University, Chung-Li District, Taoyuan, 32023, Taiwan (e-mail: jason@wavenet.cycu.edu.tw)

Abstract: In the processing industries, operating conditions often change to meet the requirements of the market and customers. To cope with the difficulty of on-line quality prediction for such multi-grade processes widely operated in process industries, a novel common and special feature extraction method is proposed for modeling multi-grade processes. A common feature extraction algorithm is proposed to determine the common directions shared by different grades of these processes. After extracting the common features, a partial least-squares modelling algorithm is used to extract the special directions for each grade, respectively. Hence, product quality prediction can be simply conducted by integrating the common and special parts of each grade for model building. A numerical case and an industrial polyethylene process are used to demonstrate the effectiveness and advantage of the proposed method.

Keywords: multi-grade process, common feature extraction, soft sensor, limited data, PLS.

1. INTRODUCTION

In many industrial plants, short time-to-market processes are often operated over a range of conditions to produce various grades of a product (e.g., a high-molecular polymer) (Jaeckle and MacGregor, 1998; 2000). Also, it is a common practice that products with different specifications but in the same category are produced by simply changing the recipe from one process to another process. In other words, these processes inherently belong to the same chemical or physical principles, although they have different operating conditions, duration time, and product sizes etc. (Lu, Yao and Gao, 2009). Take the industrial polyethylene process for example, different product grades are produced with different proportions of plastic materials and quality requirements (such as melt index (MI)) (Liu and Chen, 2013). In this work, the corresponding processes having the same mechanism but with different operating conditions and product specifications are defined as multi-grade processes. Owing to wide applications of multi-grade processes (Ohshima and Tanigaki, 2000; Kim, Lee, Han and Han, 2005; Liu, 2007). It has been increasingly appealed to develop advanced on-line quality prediction methods for these processes. Note that grade changeover is usually conducted by operator in engineering applications for operating a multi-grade process, which could result in a large settling time, overshoot, and off grade products. Besides, establishing a detailed principle model for each grade of these processes may require a long time, which could be troublesome or even impractical for application (Ohshima and Tanigaki, 2000; Kim, Lee, Han and Han, 2005).

Alternatively, process data have become widely available in many chemical plants and thus data-driven soft sensors are applied to predict product qualities that are difficult to measure on-line (Kadlec, Gabrys and Strandt, 2009). In the last few decades, data-driven soft sensors, such as partial least squares (PLS) (Burnham, Viveros & MacGregor, 1996), support vector regression (SVR) (Lee, Song, Song and Yoon, 2005) and least square support vector regression (LSSVR) (Shi and Liu, 2006), and Gaussian process regression (Ge, Chen and Song, 2011) have been widely used for process quality prediction. These methods utilize measured process data rather than a prior process knowledge. However, the traditional data-driven methods inherently restrict to the case of single population sample, and a sufficient sample data set of the process should be adopted to these methods. For multi-grade processes, it is often encountered that there is no sufficient process data to establish a prediction model for each grade of product. It is desired to develop a unified soft sensor model to predict for each grade of product in the same category. A natural idea is to extract common features from all data sets of different products for building models. The existing methods such as the 3-way factor analysis methods (Kroonenberg and Leeuw, 1980), generalized Procrustes analysis (Tenberge, 1977), generalized canonical analysis (Dahl and Naes, 2006), common PCA (CPCA) (Flury, 1984) etc. may be adopted for this purpose, constructing a common loading matrix which represents the common features for multiple-source data. A structured overview of simultaneous component methods was offered for an integrative analysis of coupled data (Van Deun, , Smilde, van der Werf, Kiers,and Van Mechelen, 2009). These methods all contributed to extending the methods from single population samples to the multiple population samples. Generally, they intended to
figure out a ‘consensus’ type of matrix which represents the common information for all data sets. However, the assumption of these methods may often have to be rejected as it is highly restrictive to a constraint that the orthogonal principal axes are the same for all data sets. In practice, the underlying distribution of each data set may differ from process to process. Note that in above mentioned methods, only the overall similarity and covariance are considered, but the individual features within each process are neglected when extracting common features. In order to conduct a comprehensive investigation into the common structure in variable correlations of multi-set data, a two-step basis vector extraction algorithm multi-set regression analysis (MsRA) is designed, which focuses on matching the pure variable correlations from one set to another (Zhao and Gao, 2012). The extracted basis vectors focus on maximizing their cross-set linear correlations. These basis vectors are regarded to be common ones and can be employed to explain the variable similarity over sets. However, due to a rank-deficient problem involved with such computation, the measured data are reduced and replaced by combination coefficients in the first step, which is hard to interpret the physical meaning and may degrade the overall prediction performance (Zhao, Yao, Gao and Wang, 2010).

In this paper, a comprehensive quality prediction model is proposed for multi-grade processes. Two steps are included in the proposed method. Firstly, the proposed common feature extraction method is applied to extract the shared features based on all the determined relevant data, which maximizes the correlation between process variables and quality variables while minimizing the scatter between different grades of processes. Secondly, by subtracting the determined common parts, special features of each grade are sequentially computed by PLS, respectively. Thus, each grade of process is divided into common, special and residual part in the proposed method. The quality variables are predicted correspondingly by the common and special part. Owing to limited data available, the proposed method shows advantages than the existing methods separately modeling different grades of processes. For clarity, the paper is organized as follows: The problem description and the proposed modelling method is detailed in Section 2. In Section 3, a numerical case and an industrial polyethylene process, are given to demonstrate the effectiveness and merit of the proposed method. Finally, some conclusions were drawn in Section 4.

2. PROPOSED MULTIPHASE PARTITION METHOD

2.1 Problem description

To cater for various market demands, a manufacturing system often produces different grades of products by simply changing material types, process operations, conditions etc., which is called multi-grade processes (Liu and Chen, 2013). Owing to grade changeover is a typical manual operation in multi-grade processes, it is critical to develop an advanced control system to provide optimal grade changeover trajectories to reduce the amount of off-grade materials. The key factor in developing such a system is a reliable and accurate model for on-line prediction of product qualities.

It should be recognized that if there are enough samples measured in each grade of process, it is good enough to establish prediction models separately. However, it is often encountered in engineering applications that the samples measured in each grade of these processes are limited or insufficient for model building with good accuracy. Suppose there are \( M \) grades in these processes. The process and quality data are denoted as \( X_i \in \mathbb{R}^{N_i \times J_i} \) and \( Y_i \in \mathbb{R}^{N_i \times J_i} \), \( i = 1, \ldots, M \), where \( J_i \) and \( J_j \) are the numbers of process variables and quality variables, respectively, that are assumed to be the same for all grades, \( N_i \) is the number of measured samples for the \( i \)th grade, \( M \) is the total number of multi-grade processes. Note that \( N_i \) may be a smaller number with respect to \( J_i \) and \( J_j \) in such application. Due to the limited data for each grade, it is less reliable or accurate to predict all processes by using the models separately established for each grade. Another difficulty is the nonlinear characteristics existed in most processes. However, most of the current published methods belong to linear methods. So it is a challenge to establish an accurate quality prediction for such processes.

There are two typical characteristics associated with multi-grade processes: 1) The samples measured in different grades share some common features; 2) The features in each process are not entirely identical with those of other processes, i.e., each process has its own special features. To the best of our knowledge, the issue of how to extract the shared common features and establish a nonlinear on-line quality prediction model with limited data for multi-grade process has not been fully researched. The main difficulty to establish an efficient prediction model for multi-grade processes lies with how to extract the common and special features of each grade. The extracted common features should ensure the best prediction performance, that is, maximize the correlation between process variables and quality variables while minimize the scatter between different grades of processes. It should be noted that the total samples of these processes, denoted by \( N = \sum_{i=1}^{M} N_i \), could be large enough in practice, which are therefore used to extract the shared features of each grade by using the proposed common feature extraction method. After extracting the common features, the special features can be determined by the traditional methods for each grade of processes, respectively. In this work, a novel common and special feature extraction method with respect to soft sensor is proposed. The details of the main method are given below.

2.2 The proposed method

To extract the common features shared by all grades of processes, a novel algorithm is firstly proposed based on all the measured data. Then PLS is used to further extract the special features in each grade of process. Thus, the quality variables are predicted by combination of the common feature and special features. The proposed common and special feature extraction method is detailed in the following.
To better extract the common features, the between scatter and within scatter in all process data should be taken into consideration comprehensively. Meanwhile, the extracted common features should have good correlation and explanations to the quality variables. To this end, a common direction is pursued, on which the projected data from different grades of processes are as consistent as possible. According to the discussion and analysis above, this kind of direction can be determined by maximizing the correlation of process data $X_i$ and quality data $Y_i$, while minimizing the scatter among the data sets. Suppose there are $M$ grades of process measurement data $X_i = [x_{i1}, \ldots, x_{iN_i}]^T \in \mathbb{R}^{N_i \times 1}$, $x_{in} \in \mathbb{R}^i$, and quality data $Y_i = [y_{i1}, \ldots, y_{iN_i}]^T \in \mathbb{R}^{N_i \times 1}$, $y_{in} \in \mathbb{R}^i$, $n = 1, \ldots, N_i$, $i = 1, \ldots, M$. All the measurement data is denoted as $X = [X_1^T, \ldots, X_M^T]^T \in \mathbb{R}^{N \times 1}$, and all quality data is denoted as $Y = [Y_1^T, \ldots, Y_M^T]^T \in \mathbb{R}^{N \times 1}$, where $N = \sum_{i=1}^{M} N_i$. Then the covariance of $X_i$ and $Y_i$ is

$$S_{w} = (X_i - \overline{X}_i)^T (Y_i - \overline{Y}_i)$$

where $\overline{X}_i$ and $\overline{Y}_i$ are the mean matrices, whose row vectors are $\overline{x}_i = \frac{1}{N_i} \sum_{n=1}^{N_i} x_{in}$ and $\overline{y}_i = \frac{1}{N_i} \sum_{n=1}^{N_i} y_{in}$, respectively. The overall covariance of all data sets is denoted as $S_w$.

$$S_w = \sum_{i=1}^{M} S_{w_i} = \sum_{i=1}^{M} (X_i - \overline{X}_i)^T (Y_i - \overline{Y}_i)$$

Similar to the traditional Fisher discriminate analysis (Chiang, Ruddell and Braatz, 2000), the between class scatter matrices $S^b_i$ and $S^b_w$ are defined.

$$S^b_i = \sum_{i=1}^{M} N_i (X_i - \overline{X}_i)(X_i - \overline{X}_i)^T$$

$$S^b_w = \sum_{i=1}^{M} S^b_i$$

where $\overline{X} = \frac{1}{M} \sum_{i=1}^{M} N_i \overline{X}_i$ and $\overline{Y} = \frac{1}{M} \sum_{i=1}^{M} N_i \overline{Y}_i$.

The common directions can be determined by maximizing the covariance of $X$ and $Y$, while minimizing the scatter between the data sets as the following optimization problem,

$$\max \quad \frac{\langle w^c \rangle^T S^b_w w^c}{\langle \langle w^c \rangle^T S^b w^c \rangle^T}$$

where the superscript $C$ is short for ‘common’, the vector $w^c$ and $v^c$ indicate the common directions of $X$ and $Y$, respectively. The above problem is equivalent to

$$\max \quad \langle w^c \rangle^T S^b_w w^c$$

s.t. $\langle w^c \rangle^T S^b w^c = 1$, $\langle v^c \rangle^T S^b w^c = 1$

Using Lagrangian function,

$$L = \langle w^c \rangle^T S^b_w w^c - 0.5 \lambda (\langle w^c \rangle^T S^b_w w^c - 1)$$

$$- 0.5 \mu (\langle v^c \rangle^T S^b_w v^c - 1)$$

Compute the partial differential in terms of $w^c$ and $v^c$, respectively, and by transforming, it’s easy to be obtained that

$$S_w v^c = \lambda S^b_w w^c$$

$$S^b_w w^c = \mu S^b_w w^c$$

Owing to $(w^c)^T S^b_w w^c = 1$ and $(v^c)^T S^b_w v^c = 1$, pre-multiply $(w^c)^T$ and $(v^c)^T$ on (8) and (9), respectively,

$$\lambda = \mu = \langle w^c \rangle^T S^b_w w^c$$

If $S^b_w$ and $S^b_w$ are invertible, then $w^c$ and $v^c$ can be expressed by each other, the following equations will be got

$$S_w (S^b_w)^{-1} S^b_w w^c = \lambda^2 S^b_w w^c$$

$$S^b_w (S^b_w)^{-1} S^b_w v^c = \mu^2 S^b_w v^c$$

By solving the eigenvalue problems in (11) and (12), the common directions $w^c$ and $v^c$ can be determined.

Suppose $R^c$ components are retained, the weight matrix, loading matrix and regression matrix are computed as $W^C = [w_{1c}^c, \ldots, w_{R^c}^c]$ , $P^C = [p_{1c}^c, \ldots, p_{R^c}^c]$ and $R^C = [r_{1c}^c, \ldots, r_{R^c}^c]$ , respectively. For each grade of process, the common parts and corresponding residual parts are expressed as

$$X^c = X \sum_{j=1}^{R^c} w_j^c (p_j^c)^T$$

$$Y^c = X \sum_{j=1}^{R^c} w_j^c (r_j^c)^T$$

$$E^c = X \prod_{j=1}^{R^c} (I - w_j^c (p_j^c)^T)$$

$$E^c = Y - Y^c$$

where $w_j^c = \prod_{i=1}^{R^c} (I - w_i^c (p_i^c)^T) w_i^c$, $I$ is an identity matrix with proper dimension.

After the common features are determined for all grades of these processes, PLS is performed on $E^c_i$ and $E^c_j$ to further extract the special features of each grade. The optimization problem of PLS is formulated as

$$\max \quad \langle E^c_i, w_j^c \rangle^T E^c_i v_j^c$$

s.t. $\langle w_j^c \rangle^T w_j^c = 1$, $\langle v_j^c \rangle^T v_j^c = 1$
where \( w^s_i \) and \( v^s_i \) are the special directions of \( E_{i,C}^s \) and \( E_{i,C}^v \), respectively. Using the Lagrangian multiplier, the optimization problem becomes an eigenvalue problem of \( (E_{i,C}^s)^T E_{i,C}^s (E_{i,C}^v)^T E_{i,C}^v w_i^s = \lambda_i^s w_i^s \), where \( \lambda_i^s \) indicates the eigenvalue for the \( i \)th grade. By using the most popular nonlinear iterative partial least squares algorithm (NIPALS), the weight matrix, loading matrix and regression matrix are determined by retaining \( R_i^s \) components, denoted by \( W_i^s = [w_{i,1}^s, \ldots, w_{i, R_i^s}] \), \( P_i^s = [p_{i,1}^s, \ldots, p_{i, R_i^s}] \), and \( R_i^s = [r_{i,1}^s, \ldots, r_{i, R_i^s}] \), respectively. Thus \( E_{i,C}^s \) and \( E_{i,C}^v \) are further divided into two orthogonal subspaces,

\[
E_{i,C}^s = X_i^s + E_i^s = E_{i,C}^s \sum_{j=1}^{R_i^s} w_{i,j}^s (p_{i,j}^s)^T + E_i^s
\]

\[
E_{i,C}^v = Y_i^v + E_i^v = E_{i,C}^v \sum_{j=1}^{R_i^s} w_{i,j}^v (r_{i,j}^v)^T + E_i^v
\]

where \( E_i^s \) and \( E_i^v \) are the overall residual matrices, \( R_i^s \) is the retained number of special directions for the \( i \)th grade, \( w_{i,j}^s = \prod_{k=1}^{j-1} (1 - w_{i,k}^s (p_{i,k}^s)^T)w_{i,j}^s \) is the \( j \)th modified special coefficient for the \( i \)th grade.

Based on above analysis, each process data can be divided into three parts:

\[
X_i = X_i^c + X_i^s + E_i^c
\]

\[
= X_i \sum_{j=1}^{R_i^c} w_{i,j}^c (p_{i,j}^c)^T + E_{i,c}^c \sum_{j=1}^{R_i^s} w_{i,j}^s (p_{i,j}^s)^T + E_i^c
\]

Correspondingly, the quality can also be divided into three parts as below.

\[
Y_i = Y_i^c + Y_i^s + E_i^c
\]

\[
= X_i \sum_{j=1}^{R_i^c} w_{i,j}^c (r_{i,j}^c)^T + E_{i,c}^c \sum_{j=1}^{R_i^s} w_{i,j}^s (r_{i,j}^s)^T + E_i^c
\]

Note that, the number of retained directions in both common and special part is important for practical use. The prediction performance in the special part is also affected by the number of retained directions in common parts. In order to obtain overall good prediction performance, brute-force or exhaustive search strategy can be applied to compute all potential combination of the retained directions in common and special parts.

To assess the prediction performance, root-mean-square error (RMSE) is considered and defined respectively as follows,

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{M} \sum_{n=1}^{N} (y_{i,n} - \hat{y}_{i,n})^2}{N}}
\]
for the use of the proposed method and MsRA. The ST S ST C
,, ,
STS STC
,, ,
STS STC
,, ,
() 1 () 0
() 1 () 0
() 1 () 0
ij ij ij
ij ij ij
ij ij ij
 ... the best results among the 
four methods. It is obvious that the proposed methods get 
smaller RMSE values than MsRA.

where \( i = 1, 2, 3, 4 \), \( j = 1, 2 \). Then four data sets are generated as below.

\[
X_i = \lambda_1 X w^C (p^C)^T + \lambda_2 X w^5 (p^5)^T + \lambda_3 X w^3 (p^3)^T \tag{25}
\]

\[
Y_i = \lambda_1 X w^C (r^C)^T + \lambda_2 X w^5 (r^5)^T + \lambda_3 X w^3 (r^3)^T \tag{26}
\]

Where \( i = 1, 2, 3, 4 \), \( \lambda_1 = 8 \), \( \lambda_2 = 2 \) and \( \lambda_3 = 1 \). The generated process data \( X \) and quality data \( Y \) are shown in Fig. 1, where the data in different set are indicated by different colors and shapes.

![Figure 2. Comparison of the extracted common direction of the third data set: a) the proposed method; b) MsRA](image)

In this case, MsRA is used to extract the common vectors in comparison with the proposed common feature extraction method. To better show the common feature extraction performance, the extracted directions extracted by the two methods for the third data sets are given in Fig. 2. The common direction extracted by the proposed method is very close to and along the similar direction as the real one as shown in Fig. 2 a). However, the common direction extracted by MsRA is closer to the short axis in the ellipse in Fig. 2 b), which is indeed not the real common direction. What’s more, the extracted direction is almost orthogonal to the real common direction. To better compare the common feature extraction performance, the correlation coefficients between the real and extracted loading and regression vectors are computed. The correlation coefficients of the proposed method are given in Table 1, where the correlation coefficient between the real and extracted loading vector is 0.9995, and the correlation coefficient between the real and extracted regression vector is 0.9973. Correspondingly, the correlation coefficients of the MsRA for different data sets are given in Table 2. Except the correlation coefficient between the real and extracted loading vector in S2, which is 0.9998 and feebly higher than 0.9995, the other correlation coefficients of MsRA are lower than that of the proposed method obviously.

<table>
<thead>
<tr>
<th>Data set 1</th>
<th>Data set 2</th>
<th>Data set 3</th>
<th>Data set 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loading vector</td>
<td>-0.8788</td>
<td>0.9998</td>
<td>-0.9391</td>
</tr>
<tr>
<td>Regression vector</td>
<td>-0.9290</td>
<td>0.9625</td>
<td>-0.8363</td>
</tr>
</tbody>
</table>

### 3.2 Polyethylene process

Consider an industrial polyethylene process studied in the reference (Liu and Chen, 2013) for quality prediction on MI. In this study, all the samples have been collected from daily process records and the corresponding laboratory analysis. For steady-state grades, the product quality of the polyethylene production was sampled and analyzed in the lab once a day, always in the morning. When the production changes its grade, the time spent for the lab analysis on each sample was about 4–6 hours. Owing to the lack of on-line analyzer, the manipulated variables could not be conducted until the off-line assay results were obtained. Consequently, off-grade products and materials had been undesirably produced to a certain amount. Here, steady-state grades, denoted as S1, S2, S3, are considered for analysis.

In this case, MsRA and the proposed method are applied to predict MI in the polyethylene process. The number of retained common and special directions are \( R^C = 1 \) and \( R^S = 4 \) for the use of the proposed method and MsRA. The RMSE computed by the four methods are given in Table 3, where the red numbers indicate the best results among the four methods. It is obvious that the proposed methods get smaller RMSE values than MsRA.
Table 3. RMSE for each stable data set in the polyethylene process

<table>
<thead>
<tr>
<th></th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed method</td>
<td>0.3433</td>
<td>0.9070</td>
<td>0.7612</td>
</tr>
<tr>
<td>MsRA</td>
<td>0.3723</td>
<td>0.6754</td>
<td>0.7759</td>
</tr>
</tbody>
</table>

4. CONCLUSIONS

In this paper, a novel quality prediction method based on common and special feature extraction has been proposed for industrial multi-grade processes. An important merit is that not only the common features shared by all grades are efficiently extracted, but also the special features of each grade are sequentially extracted for model building. Based on the common and special feature extraction, quality prediction can be conveniently conducted by integrating the common and special parts of each grade. A numerical case has been used to demonstrate the advantage of the proposed method for common feature extraction in comparison with the existing methods such as MsRA. The application to an industrial polyethylene process based on collected samples from the lab assay has well demonstrated the prediction performance of the proposed method compared to the existing methods. It should be noted that the advantages of proposed method is evident when the samples for each grade is very limited. However, when there are sufficient numbers of samples for each grade, the traditional PLS modeling method could provide the similar prediction performance with the proposed method. It therefore deserves further exploration on the number of data for a suitable choice of these methods in the future.

ACKNOWLEDGEMENT

This work is supported in part by the NSF China Grants 61473054 and 61633006, and the Fundamental Research Funds for the Central Universities of China.

REFERENCES


