基于张量偏最小二乘法的高维输出预测模型

冯 宇
(长安大学 电子与控制工程学院,陕西 西安 710064)

摘 要:针对输出为高维数据的预测问题,构建了一种基于张量偏最小二乘法的预测模型。该模型的输入和输出均为张量,可在不进行数据降维操作的前提下对多个输出同时预测。该方法以偏最小二乘法为基础,引入逐块 Tucker 分解和高阶奇异值分解,通过对输入和输出变量的分析,提取同时包含二者最大信息的潜在变量并计算残差,再通过残差计算新的潜在变量,循环直到残差小于给定范围为止。实验数据来源于心脏传导系统在正常和急性高糖环境下采集的电生理信息,通过最大正振幅、最大负振幅、频率、单次信号持续时间四个维度的输入同时预测急性高糖的浓度和作用时间,并将预测结果与传统的多向偏最小二乘法和多维偏最小二乘法相比较。实验结果表明,基于张量偏最小二乘法的预测模型预测精度最高。

关键词:预测模型;高维输出;张量;偏最小二乘法;心脏电生理信息
中图分类号:TP399 **文献标识码**:A **文章编号**:1673-629X(2019)07-0114-05
doi:10.3969/j.issn.1673-629X.2019.07.023

A Multi-dimensional Output Prediction Model Based on Tensor Partial Least Squares

FENG Yu
(School of Electronic and Control Engineering, Chang'an University, Xi'an 710064, China)

Abstract: A prediction model based on tensor partial least squares (TPLS) is built to solve the problem of high dimensional data. The input and output of the model are tensors, and multi-dimensional output can be predicted without dimensional reduction. Based on partial least squares, this method introduces block by block Tucker decomposition and high order singular value decomposition. Through the analysis of input and output variables, the potential variables containing the maximum information of both are extracted and the residual error is calculated. Then, the new potential variable is calculated through the residual error, and the loop runs until the residual error is less than the given range. The experimental data are derived from electrophysiological information of cardiac conduction systems in normal and acute hyperglycemic environments. The input is a four-dimensional tensor which contains maximum positive amplitude, maximum negative amplitude, frequency and single signal duration. The predictive output contains the concentrations and durations of acute hyperglycemia. The prediction results are compared with multi-way PLS and N-way PLS, which shows that the tensor PLS has the highest prediction accuracy.

Key words: prediction model; multi-dimensional output; tensor; PLS; cardiac electrophysiological information

0 引 言

预测模型广泛应用于科学研究和工业生产的各个领域,系统的预测模型是指用数学语言或公式来描述系统的输入与输出间的关系,其主要功能是建立连续或离散的函数模型,预测给定自变量对应的因变量的值[1],其数学本质是建立一个预测函数,使得对于每一个提前期,实际值与预测值之间的偏差的均方尽可能小[2]。建立预测模型有多种思路,例如趋势外推预测、回归预测、卡尔曼滤波预测、组合预测等[3-6],在这些思路的基础上已有多种预测模型被建立并广泛使用,但多数预测模型针对的是一维输出系统,即需要预测的参数是一维的。随着数据量和系统复杂程度的增加,需要寻求建立同时预测多维输出的预测模型。文中研究基于张量偏最小二乘法(tensor partial least squares, TPLS)的高维输入输出预测模型的建模方法,该方法可在不进行降维操作的情况下直接处理输入输

收稿日期:2018-09-17 **修回日期**:2019-01-23 **网络出版时间**:2019-03-21
基金项目:陕西省自然科学基础研究计划(2017JQ6075);中央高校基本科研业务费专项资金(300102328103)
作者简介:冯 宇(1984-),男,博士,讲师,研究方向为数据挖掘、生物信号测量与分析。
网络出版地址:http://kns.cnki.net/kcms/detail/61.1450.TP.20190321.0942.062.html

出均为张量的高维数据,从而降低了因数据结构遭到破坏而导致信息丢失的风险。建模实验数据来源于心脏传导系统的电生理信息,通过正常和急性高血糖环境下电生理信息数据的特征,同时预测急性高血糖的浓度和作用时间。

1　关键技术

1.1　相关定义与符号

文中使用下划线黑体大写字母表示 N 维张量($n \geqslant 3$),例如 $\underline{\boldsymbol{X}}$;用黑体大写字母表示矩阵,例如 $\boldsymbol{Y}$;用黑体小写字母表示向量,例如 $\boldsymbol{v}$;用斜体小写字母表示标量,例如 a;用 v_i 表示向量 $\boldsymbol{v}$ 的第 i 个分量;用 y_{ij} 表示矩阵 $\boldsymbol{Y}$ 的第 (i,j) 个元素;用 $x_{i_1,i_2,\cdots,i_N}$ 或 $(\underline{\boldsymbol{X}})_{i_1,i_2,\cdots,i_N}$ 表示 N 维张量的第 $(i_1,i_2,\cdots,i_N)$ 个元素($\underline{\boldsymbol{X}} \in \mathrm{IR}^{I_1 \times I_2 \times \cdots \times I_N}, i_N = 1,2,\cdots,I_N$)。用 $\boldsymbol{A}^{(n)}$ 表示第 n 个系数矩阵。用符号“×”表示 n 模乘(n-mode product)[7-8],张量 $\underline{\boldsymbol{X}} \in \mathrm{IR}^{I_1 \times I_2 \times \cdots \times I_N}$ 和矩阵 $\boldsymbol{A} \in \mathrm{IR}^{J_n \times I_n}$ 的 n 模乘可以写为:

$$\underline{\boldsymbol{Y}} = \underline{\boldsymbol{X}} \times_n \boldsymbol{A} \in \mathrm{IR}^{I_1 \times \cdots \times I_{n-1} \times J_n \times I_{n+1} \times \cdots \times I_N} \tag{1}$$

按照元素可以表示为:

$$y_{i_1 i_2 \cdots i_{n-1} j i_{n+1} \cdots i_N} = \sum_{i_n} x_{i_1 i_2 \cdots i_n \cdots i_N} a_{j i_n} \tag{2}$$

1.2　张量偏最小二乘法的分解与优化

可以认为张量偏最小二乘法是经典最小二乘法在高维上的拓展,其思路是在高阶偏最小二乘法(high order partial least squares,HOPLS)的基础上考虑了高维数据间的相关程度问题。图 1 为张量偏最小二乘法的分解示意图[8]。将张量 $\underline{\boldsymbol{X}}$ 分解为一组秩-$(1,L_1,L_2,\cdots,L_N)$ 之和,张量 $\underline{\boldsymbol{Y}}$ 以满足与 $\underline{\boldsymbol{X}}$ 存在共同的潜在元素 $\boldsymbol{T}$ 为条件进行分解。

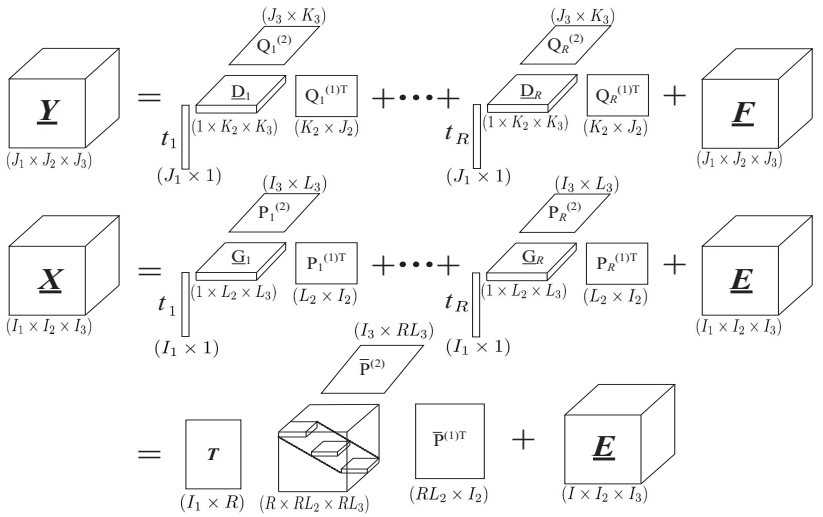

图 1　张量偏最小二乘法分解示意

在分解过程中引入逐块 Tucker 分解方法[9-10],即将张量 $\underline{\boldsymbol{X}}$ 和张量 $\underline{\boldsymbol{Y}}$ 均分解为一组秩-$(1,L_2,\cdots,L_N)$ 的 Tucker 块,两组分解包含相同的潜在元素。图 1 中的分解可以表示为:

$$\begin{cases} \underline{\boldsymbol{X}} = \sum_{r=1}^{R} \underline{\boldsymbol{G}}_r \times_1 \boldsymbol{t}_r \times_2 \boldsymbol{P}_r^{(1)} \times_3 \cdots \times_N \boldsymbol{P}_r^{(N-1)} + \underline{\boldsymbol{E}}_R \\ \underline{\boldsymbol{Y}} = \sum_{r=1}^{R} \underline{\boldsymbol{D}}_r \times_1 \boldsymbol{t}_r \times_2 \boldsymbol{Q}_r^{(1)} \times_3 \cdots \times_M \boldsymbol{Q}_r^{(N-1)} + \underline{\boldsymbol{F}}_R \end{cases} \tag{3}$$

其中,R 为隐向量个数;$\boldsymbol{t}_r \in \mathrm{IR}^{I_1}$ 为第 r 个隐向量;$\{\boldsymbol{P}_r^{(n)}\}_{n=1}^{N-1} \in \mathrm{IR}^{I_{n+1} \times L_{n+1}}$ 为 mode-n 的负载矩阵;$\{\boldsymbol{Q}_r^{(m)}\}_{m=1}^{M-1} \in \mathrm{IR}^{J_{m+1} \times K_{m+1}}$ 为 mode-m 的负载矩阵;$\underline{\boldsymbol{G}}_r \in \mathrm{IR}^{1 \times L_2 \times \cdots \times L_N}$ 和 $\underline{\boldsymbol{D}}_r \in \mathrm{IR}^{1 \times K_2 \times \cdots \times K_N}$ 为核张量。

若考虑模型输入参数之间的相关性,还应引入相关性张量,可将该张量与式 3 中的 $\underline{\boldsymbol{E}}_R$ 相加得到 $\underline{\boldsymbol{E}}_R'$,用 $\underline{\boldsymbol{E}}_R'$ 代替式中的 $\underline{\boldsymbol{E}}_R$。

为了解决唯一性问题和潜变量的维数问题,可以增加约束条件:

$$\boldsymbol{P}_r^{(n)\mathrm{T}} \boldsymbol{P}_r^{(n)} = \boldsymbol{I}, \quad \boldsymbol{Q}_r^{(m)\mathrm{T}} \boldsymbol{Q}_r^{(m)} = \boldsymbol{I} \tag{4}$$

$$\| \boldsymbol{t}_r \|_F = 1 \tag{5}$$

式 4 表示负载矩阵序列是列正交的;

式 5 表示潜变量的维数为 1。

定义隐向量 $\boldsymbol{T} = [t_1, t_2, \cdots, t_R]$,mode-n 对应的负载矩阵 $\bar{\boldsymbol{P}}^{(n)} = [\boldsymbol{P}_1^{(n)}, \boldsymbol{P}_2^{(n)}, \cdots, \boldsymbol{P}_R^{(n)}]$,mode-m 对应的负载矩阵 $\bar{\boldsymbol{Q}}^{(m)} = [\boldsymbol{Q}_1^{(m)}, \boldsymbol{Q}_2^{(m)}, \cdots, \boldsymbol{Q}_R^{(m)}]$;

定义核张量:

$$\begin{cases} \bar{\underline{\boldsymbol{G}}} = \mathrm{blockdiag}(\underline{\boldsymbol{G}}_1, \underline{\boldsymbol{G}}_2, \cdots, \underline{\boldsymbol{G}}_R) \in \mathrm{IR}^{R \times RL_2 \times \cdots \times RL_N} \\ \bar{\underline{\boldsymbol{D}}} = \mathrm{blockdiag}(\underline{\boldsymbol{D}}_1, \underline{\boldsymbol{D}}_2, \cdots, \underline{\boldsymbol{D}}_R) \in \mathrm{IR}^{R \times RK_2 \times \cdots \times RK_N} \end{cases} \tag{6}$$

则式 3 可以写成:

$$\begin{cases} \underline{\boldsymbol{X}} = \underline{\overline{\boldsymbol{G}}} \times_1 \boldsymbol{T} \times_2 \overline{\boldsymbol{P}}^{(1)} \times_3 \cdots \times_N \overline{\boldsymbol{P}}^{(N-1)} + \underline{\boldsymbol{E}}_R \\ \underline{\boldsymbol{Y}} = \underline{\overline{\boldsymbol{D}}} \times_1 \boldsymbol{T} \times_2 \overline{\boldsymbol{Q}}^{(1)} \times_3 \cdots \times_M \overline{\boldsymbol{Q}}^{(M-1)} + \underline{\boldsymbol{F}}_R \end{cases} \tag{7}$$

其中，$\underline{\boldsymbol{E}}_R$ 和 $\underline{\boldsymbol{F}}_R$ 为残差。

从图 1 可以看出，核张量 $\underline{\overline{\boldsymbol{G}}}$ 和 $\underline{\overline{\boldsymbol{D}}}$ 具有特殊的对角块结构。

通常情况下，可以使用顺序法来提取潜在元素，即先提取一个元素，通过该元素的收缩张量计算出残差，再通过该残差计算另一个元素。式 3 的分解过程可以看作一个优化问题：将 $\underline{\boldsymbol{X}}$ 和 $\underline{\boldsymbol{Y}}$ 近似成正交 Tucker 模型，该模型对给定的 mode 有相同的潜在元素。若分别对 $\underline{\boldsymbol{X}}$ 和 $\underline{\boldsymbol{Y}}$ 进行高阶奇异值分解（high order singular value decomposition，HOSVD）[11]，可以得到 $\underline{\boldsymbol{X}}$ 的最佳近似秩-$(1, L_2, \cdots, L_N)$ 和 $\underline{\boldsymbol{Y}}$ 的最佳近似秩-$(1, K_2, \cdots, K_M)$，但却不一定能得到共同隐向量 t_r。另一种方法是先计算出 $\underline{\boldsymbol{X}}$ 的最佳近似，然后通过已知的 t_r 来近似得到 $\underline{\boldsymbol{Y}}$，这种方法的缺点是共同潜变量不一定可以很好地预测 $\underline{\boldsymbol{Y}}$。

根据式 3，子空间变换的优化问题可以转换成另一个问题：确定正交负载矩阵 $\boldsymbol{P}_r^{(n)}$、$\boldsymbol{Q}_r^{(m)}$，$(r = 1, 2, \cdots, R)$ 和对应的 $\boldsymbol{t}_r$，以满足一定的准则。因为每一项都可以按照相同准则逐个优化，之后的工作就是找到第一个隐向量 $\boldsymbol{t}_r$ 和负载矩阵 $\boldsymbol{P}^{(n)}$ 和 $\boldsymbol{Q}^{(m)}$。

为了在保证共同隐向量 $\boldsymbol{t}$ 的前提下使残差 $\boldsymbol{E}$ 和 $\boldsymbol{F}$ 最小，需要引入三个定理：

定理 1：对于给定的张量 $\underline{\boldsymbol{X}} \in \mathrm{IR}^{I_1 \times \cdots \times I_N}$ 和列正交矩阵 $\boldsymbol{P}^{(n)} \in \mathrm{IR}^{I_{n+1} \times L_{n+1}}$ $(n = 1, 2, \cdots, N-1)$ 以及向量 $\boldsymbol{t} \in \mathrm{IR}^{I_1}$（$\| \boldsymbol{t} \|_F = 1$），有：

$$\min_{\underline{\boldsymbol{G}}} \| \underline{\boldsymbol{X}} - \underline{\boldsymbol{G}} \times_1 t \times_2 \boldsymbol{P}^{(1)} \times_3 \cdots \times_N \boldsymbol{P}^{(N-1)} \|_F^2 \tag{8}$$

式 8 的最小二乘解可以写为：

$$\underline{\boldsymbol{G}} = \underline{\boldsymbol{X}} \times_1 \boldsymbol{t}^{\mathrm{T}} \times_2 \boldsymbol{P}^{(1)\mathrm{T}} \times_3 \cdots \times_N \boldsymbol{P}^{(N-1)\mathrm{T}} \tag{9}$$

定理 2：对于给定张量 $\underline{\boldsymbol{X}} \in \mathrm{IR}^{I_1 \times \cdots \times I_N}$，以下的两个约束优化问题等价：

$$\min_{\{\boldsymbol{P}^{(n)}, \boldsymbol{t}, \underline{\boldsymbol{G}}\}} \| \underline{\boldsymbol{X}} - \underline{\boldsymbol{G}} \times_1 \boldsymbol{t} \times_2 \boldsymbol{P}^{(1)} \times_3 \cdots \times_N \boldsymbol{P}^{(N-1)} \|_F^2$$
$$\text{s.t. 矩阵 } \boldsymbol{P}^{(n)} \text{ 列正交，且 } \| \boldsymbol{t} \|_F = 1 \tag{10}$$

$$\max_{\{\boldsymbol{P}^{(n)}, \boldsymbol{t}\}} \| \underline{\boldsymbol{X}} \times_1 \boldsymbol{t}^T \times_2 \boldsymbol{P}^{(1)T} \times_3 \cdots \times_N \boldsymbol{P}^{(N-1)T} \|_F^2 \tag{11}$$
$$\text{s.t. 矩阵 } \boldsymbol{P}^{(n)} \text{ 列正交，且 } \| t \|_F = 1$$

定理 3：若 $\underline{\boldsymbol{G}} \in \mathrm{IR}^{1 \times L_2 \times \cdots \times L_N}$ 且 $\underline{\boldsymbol{K}} \in \mathrm{IR}^{1 \times K_2 \times \cdots \times K_M}$，则

$$\| < \underline{\boldsymbol{G}}, \underline{\boldsymbol{D}} >_{\{1;1\}} \|_F^2 = \| \underline{\boldsymbol{G}} \|_F^2 \cdot \| \underline{\boldsymbol{D}} \|_F^2 \tag{12}$$

三个定理的证明可参考文献[8-9,12]。

假设已知正交矩阵 $\boldsymbol{P}^{(n)}$ 和 $\boldsymbol{Q}^{(m)}$ 以及向量 $\boldsymbol{t}$，根据定理 1，式 3 的核张量可以写成如下形式：

$$\begin{cases} \underline{\boldsymbol{G}} = \underline{\boldsymbol{X}} \times_1 \boldsymbol{t}^T \times_2 \boldsymbol{P}^{(1)T} \times_3 \cdots \times_N \boldsymbol{P}^{(N-1)T} \\ \underline{\boldsymbol{D}} = \underline{\boldsymbol{Y}} \times_1 \boldsymbol{t}^T \times_2 \boldsymbol{Q}^{(1)T} \times_3 \cdots \times_M \boldsymbol{Q}^{(M-1)T} \end{cases} \tag{13}$$

根据定理 2，最小化 $\| \underline{\boldsymbol{E}} \|_F$ 和 $\| \underline{\boldsymbol{F}} \|_F$ 等价于最大化 $\| \underline{\boldsymbol{G}} \|_F$ 和 $\| \underline{\boldsymbol{D}} \|_F$。但是目前没有一种张量分解方法可以根据 $\{\boldsymbol{P}^{(n)}\}_{n=1}^{N-1}$ 和 $\{\boldsymbol{Q}^{(m)}\}_{m=1}^{M-1}$ 以及 $\boldsymbol{t}$ 同时最大化 $\| \underline{\boldsymbol{G}} \|_F$ 和 $\| \underline{\boldsymbol{D}} \|_F$，因此，需要引入核张量积的范数，并进行最大化，即 $\max\{ \| \underline{\boldsymbol{G}} \|_F^2 \cdot \| \underline{\boldsymbol{D}} \|_F^2 \}$。由于隐向量 $\boldsymbol{t}$ 是由 $\boldsymbol{P}^{(n)}$ 和 $\boldsymbol{Q}^{(m)}$ 确定的，所以首先应优化正交负载，然后通过正交负载计算隐向量。

根据定理 3，最大化 $\| \underline{\boldsymbol{G}} \|_F^2 \cdot \| \underline{\boldsymbol{D}} \|_F^2$ 等价于最大化 $\| < \underline{\boldsymbol{G}}, \underline{\boldsymbol{D}} >_{\{1;1\}} \|_F^2$，又根据式 13 和 $\boldsymbol{t}^{\mathrm{T}}\boldsymbol{t} = 1$，$\| < \underline{\boldsymbol{G}}, \underline{\boldsymbol{D}} >_{\{1;1\}} \|_F^2$ 可以表示为：

$$\| < \underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}} >_{\{1;1\}} ; \boldsymbol{P}^{(1)\mathrm{T}} \cdots \boldsymbol{P}^{(N-1)\mathrm{T}} \boldsymbol{Q}^{(1)\mathrm{T}} \cdots \boldsymbol{Q}^{(M-1)\mathrm{T}} \|_F^2 \tag{14}$$

式 14 的形式与二维偏最小二乘法的优化问题很相似，式中的交叉协方差矩阵 $\boldsymbol{X}^{\mathrm{T}}Y$ 被 $< \underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}} >_{\{1;1\}}$ 所替换。若定义 $< \underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}} >_{\{1;1\}}$ 为一个 mode-1 的交叉协方差张量 $\underline{\boldsymbol{C}}$：

$$\underline{\boldsymbol{C}} = \mathrm{COV}_{\{1;1\}}(\underline{\boldsymbol{X}}, \underline{\boldsymbol{Y}}) \in \mathrm{IR}^{I_2 \times \cdots \times I_N \times J_2 \times \cdots \times J_M} \tag{15}$$

则优化问题可以表示为：

$$\max_{\{\boldsymbol{P}^{(n)}, \boldsymbol{Q}^{(m)}\}} \| \underline{\boldsymbol{C}} \times_1 \boldsymbol{P}^{(1)\mathrm{T}} \times_2 \cdots \times_{N-1} \boldsymbol{P}^{(N-1)\mathrm{T}} \times_N \boldsymbol{Q}^{(1)\mathrm{T}} \times_{N+1} \cdots \times_{N+M-2} \boldsymbol{Q}^{(M-1)\mathrm{T}} \|_F^2 \tag{16}$$
$$\text{s.t. } \boldsymbol{P}^{(n)\mathrm{T}} \boldsymbol{P}^{(n)} = \boldsymbol{I}_{L_{n+1}}, \boldsymbol{Q}^{(m)\mathrm{T}} \boldsymbol{Q}^{(m)} = \boldsymbol{I}_{K_{m+1}}$$

其中，$\boldsymbol{P}^{(n)}(n = 1, 2, \cdots, N-1)$，$\boldsymbol{Q}^{(m)}(m = 1, 2, \cdots, M-1)$ 为待优化的参数。

根据定理 2 和 $\boldsymbol{P}^{(n)}$、$\boldsymbol{Q}^{(m)}$ 的正交性，式 16 的优化问题等价于找到张量 $\underline{\boldsymbol{C}}$ 的最佳子空间近似值：

$$\underline{\boldsymbol{C}} \approx \underline{\boldsymbol{G}}^{(C)} \times_1 \boldsymbol{P}^{(1)} \times_2 \cdots \times_{N-1} \boldsymbol{P}^{(N-1)} \times_N \boldsymbol{Q}^{(1)} \times_{N+1} \cdots \times_{N+M-2} \boldsymbol{Q}^{(M-1)} \tag{17}$$

这个值可以通过对张量 $\underline{\boldsymbol{C}}$ 的秩-$(L_2, \cdots, L_N, K_2, \cdots, K_M)$ 做高阶奇异值分解得到。

根据定理 1，式 16 中的优化项等价于核张量 $\underline{\boldsymbol{G}}^{(C)}$ 的范数。为此，使用高阶正交迭代（high order orthogonal iteration，HOOI）算法[9,13]，通过对张量 $\underline{\boldsymbol{C}}$ 做正交 Tucker 分解来寻找 $\boldsymbol{P}^{(n)}$ 和 $\boldsymbol{Q}^{(m)}$。

据此，共同隐向量 t 便可以根据 $\boldsymbol{P}^{(n)}$ 和 $\boldsymbol{Q}^{(m)}$ 计算。由于预测模型的最终目标是根据张量 $\underline{\boldsymbol{X}}$ 预测张量 $\underline{\boldsymbol{Y}}$，与偏最小二乘法类似，需要通过 $\underline{\boldsymbol{X}}$ 计算 $\boldsymbol{t}$，同时还需要计算回归系数 $\underline{\boldsymbol{D}}$。对于已知的负载矩阵集合 $\{\boldsymbol{P}^{(n)}\}$，隐向量 $\boldsymbol{t}$ 应对张量 $\underline{\boldsymbol{X}}$ 的方差具有最大的解释能力，即：

$$\boldsymbol{t} = \arg\min_{\boldsymbol{t}} \| \underline{\boldsymbol{X}} - \underline{\boldsymbol{G}} \times_1 \boldsymbol{t} \times_2 \boldsymbol{P}^{(1)} \times_2 \cdots \times_N \boldsymbol{P}^{(N-1)} \|_F^2 \tag{18}$$

式 18 可通过高阶正交迭代得到,只要选择 $\boldsymbol{t}$ 为矩阵 $(\underline{\boldsymbol{X}} \times_2 \boldsymbol{P}^{(1)\mathrm{T}} \times_3 \cdots \times_N \boldsymbol{P}^{(N-1)\mathrm{T}})_{(1)}$ 的最大左奇异矩阵即可[14]。核张量 $\underline{\boldsymbol{G}}$ 和 $\underline{\boldsymbol{D}}$ 可由式 13 计算得到。

上述的计算步骤循环进行,直到所有元素都被计算一次或者残差小于给定的范围为止。

1.3　张量偏最小二乘法算法描述

输入:$\underline{\boldsymbol{X}} \in \mathrm{IR}^{J_1 \times \cdots \times J_N}$,$\underline{\boldsymbol{Y}} \in \mathrm{IR}^{J_1 \times \cdots \times J_M}$;

// $I_1 = J_1$;特征向量的个数为 R,负载向量的个数为 $\{L_n\}_{n=2}^{N}$ 和 $\{K_m\}_{m=2}^{M}$

输出:$\{\boldsymbol{P}_r^{(n)}\}$,$\{\boldsymbol{Q}_r^{(m)}\}$,$\{\underline{\boldsymbol{G}}_r\}$,$\{\underline{\boldsymbol{D}}_r\}$,$\boldsymbol{T}$;

// $r = 1,2,\cdots,R$;$n = 1,2,\cdots,N-1$;$m = 1,2,\cdots,M-1$

$\underline{\boldsymbol{E}}_1 = \underline{\boldsymbol{X}}$,$\underline{\boldsymbol{F}}_1 = \underline{\boldsymbol{Y}}$;//初始化

for $r = 1$ to R do

{

if($\| \underline{\boldsymbol{E}}_r \|_F > \varepsilon$ and $\| \underline{\boldsymbol{F}}_r \|_F > \varepsilon$)

{

$\underline{\boldsymbol{C}}_r = < \underline{\boldsymbol{E}}_r, \underline{\boldsymbol{F}}_r >_{\{1,1\}}$;

$\underline{\boldsymbol{C}}_r \approx \underline{\boldsymbol{G}}^{(C_r)} \times_1 \boldsymbol{P}_r^{(1)} \times_2 \cdots \times_{N-1} \boldsymbol{P}_r^{(N-1)} \times_N \boldsymbol{Q}_r^{(1)} \times_{N+1} \cdots \times_{N+M-2} \boldsymbol{Q}_r^{(M-1)}$;

// $\underline{\boldsymbol{C}}_r$ 的 Rank - $(L_2,\cdots,L_N,K_2,\cdots,K_M)$ 正交 Tucker 分解

$\boldsymbol{t}_r$ = 最大左奇异矩阵 SVD[$(\underline{\boldsymbol{E}}_r \times_2 \boldsymbol{P}_r^{(1)\mathrm{T}} \times_3 \cdots \times_N \boldsymbol{P}_r^{(N-1)\mathrm{T}})_{(1)}$];

$\underline{\boldsymbol{G}}_r = \underline{\boldsymbol{E}}_r \times_1 \boldsymbol{t}_r^{\mathrm{T}} \times_2 \boldsymbol{P}_r^{(1)\mathrm{T}} \times_3 \cdots \times_N \boldsymbol{P}_r^{(N-1)\mathrm{T}}$;

$\underline{\boldsymbol{D}}_r = \underline{\boldsymbol{F}}_r \times_1 \boldsymbol{t}_r^{\mathrm{T}} \times_2 \boldsymbol{Q}_r^{(1)\mathrm{T}} \times_3 \cdots \times_M \boldsymbol{Q}_r^{(M-1)\mathrm{T}}$;

$\underline{\boldsymbol{E}}_{r+1} = \underline{\boldsymbol{E}}_r - (\underline{\boldsymbol{G}}_r \times_1 \boldsymbol{t}_r \times_2 \boldsymbol{P}_r^{(1)} \times_3 \cdots \times_N \boldsymbol{P}_r^{(N-1)})$;

$\underline{\boldsymbol{F}}_{r+1} = \underline{\boldsymbol{F}}_r - (\underline{\boldsymbol{D}}_r \times_1 \boldsymbol{t}_r \times_2 \boldsymbol{Q}_r^{(1)} \times_3 \cdots \times_M \boldsymbol{Q}_r^{(M-1)})$;

}

else

break;

}

2　数据来源与实验设计

本实验使用 MED64 微电极阵列测量系统(Alpha-MedScience,日本)来进行数据采集和分析。系统包括传感器、信号放大电路、控制器、采集和处理数据所用的计算机以及生物实验相关设备(显微镜、灌流槽、蠕动泵等)。实验动物使用雄性 C57/BL6J 小鼠(年龄 8 ~ 12周,体重 20 ~ 25 g),实验操作过程符合长安大学生物实验操作规程和伦理学要求。实验中首先取出小鼠心脏,使用有钙台式液和 Langendorff 离体心脏灌流方法进行离体灌流,然后将右心房置于下方,直接与传感器测量平面接触,进行信号测量。整个测量过程中,给样本持续提供有钙台式液(溶液中加入 5% 二氧化碳和 95% 氧气),流速为 5 ml/min,每次实验中给样本加入不同浓度的高糖溶液(浓度分别为 20 mM、30 mM、40 mM 和 50 mM)。在每次实验中,首先记录对照样的测量信号,随后,从第 0 分钟开始加入高糖溶液,持续加入 40 分钟。每种浓度加入后,从第 0 分钟开始,每间隔 5 分钟测量一次,采样频率为 20 kHz,每次测量持续 30 s。

在数据处理过程中,首先提取出每次测量到电位信号的最大正振幅、最大负振幅、频率、单次信号持续时间等特征值。再将提取出的特征值作为预测模型的输入矩阵,高糖作用时间和高糖浓度作为预测模型的二维输出,把实验数据分为训练集和验证集[15],训练集用于确定预测模型的系数,验证集用于检验预测结果。由于预测模型的输出矩阵的两个维度数据的量纲不同,为了准确描述二维输出的误差,需要对作用时间和溶液浓度这两个物理量进行量纲和数据的标准化。对于高糖作用时间,规定加入高糖溶液的第 0 分钟为时间轴的原点,每 5 分钟为一个单位;对于高糖溶液浓度,规定 0 mM 的高糖浓度为浓度轴的原点,每 10 mM 为一个单位。实验中选用预测均方根误差(RMSEP)、交叉验证均方根误差(RMSECV)、预测平方相关系数(R_{p}^2)和交叉验证平方相关系数(R_{cv}^2)几个参数来检验预测模型的预测能力。同时,对于相同的数据,实验中选用多向偏最小二乘法(MPLS)和多维偏最小二乘法(NPLS)进行建模[16-18],以便比较几种模型的预测效果。因为预测输出是二维数据,计算误差的方法是计算预测值和实际值在作用时间-溶液浓度平面上的欧氏距离,即:

$$E_n = \sqrt{(t_n - t_n^*)^2 + (C_n - C_n^*)^2} \tag{19}$$

其中,t_n 为第 n 个实际输出值的时间分量;t_n^* 为第 n 个预测输出值的时间分量;C_n 为第 n 个实际输出值的浓度分量;C_n^* 为第 n 个预测输出值的浓度分量。

3　实验结果与分析

表 1 给出了几种方法的输出结果,图 2 给出了表 1 中三种模型输出的预测值和实际值的对比。可以看出,使用预测值和实际值在作用时间-溶液浓度平面上的欧氏距离作为评价标准的情况下,张量偏最小二乘法的预测结果明显好于另外两种方法,其预测均方根误差(root-mean-squared error of prediction,RMSEP)分别比多向偏最小二乘法和多维偏最小二乘

表 1　三种二维输出预测模型的预测结果

误差	RMSEP	RMSECV	R_{p}^2	R_{cv}^2
MPLS	0.563 2	0.553 8	0.643 4	0.646 6
NPLS	0.600 9	0.546 6	0.587 7	0.667 9
TPLS	0.374 4	0.390 7	0.876 0	0.806 9

法的预测均方根误差小 33.5% 和 37.7%；交叉验证均方根误差（RMSECV，root-mean-square error of cross validation）分别比多向偏最小二乘法和多维偏最小二乘法的交叉验证均方根误差小了 29.4% 和 28.5%。在三种方法中，张量偏最小二乘法的预测平方相关系数（R_p^2）和交叉验证平方相关系数（R_{cv}^2）也都是最大的，说明该方法的预测输出的离散程度最小。

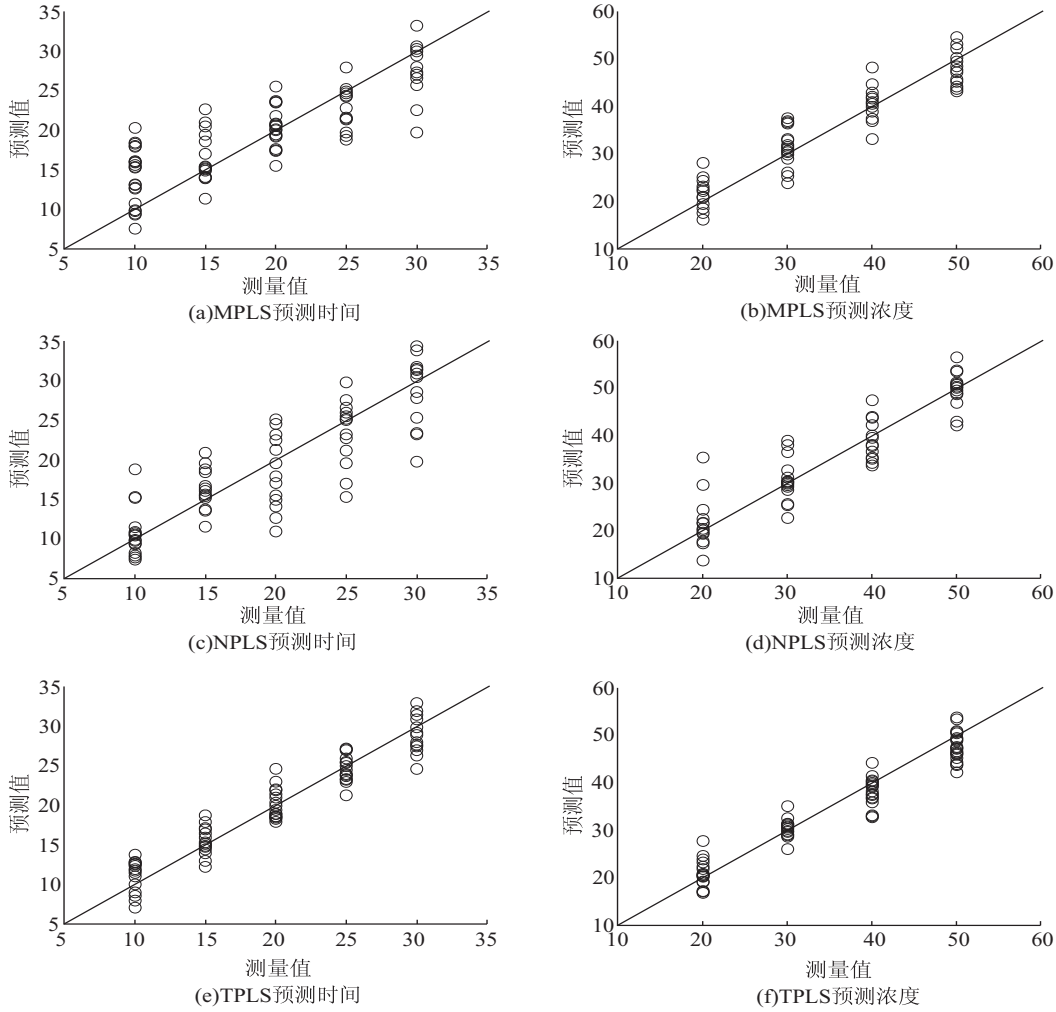

图 2 三种方法预测值和实际值的对比

4 结束语

文中使用基于张量偏最小二乘法的高维输入输出预测模型，实现了高糖溶液对生物样本的作用时间和高糖溶液浓度两个输出变量的同时预测。与传统的多向偏最小二乘法和多维偏最小二乘法相比较，基于张量偏最小二乘法的预测模型最适合用于对急性高血糖浓度和作用时间同时预测。

未来应研究不同维度的输出数据间存在关联时预测模型的优化问题，并且应关注该方法的临床可操作性，即如何在体表提取与急性高血糖密切相关的电生理信号，使该方法可以为急性高血糖的临床医学诊断提供指导与建议。

参考文献：

[1] HAN J，KAMBER M，PEI J. 数据挖掘概念与技术[M]. 范 明，孟小峰，译. 北京：机械工业出版社，2012.

[2] BOX G，JENKINS G，REINSEL G. 时间序列分析预测与控制[M]. 王成璋，尤梅芳，郝 杨，译. 北京：机械工业出版社，2011.

[3] 夏昌浩，曹 瑾，张 密，等. 电力负荷趋势外推预测算例分析与模型检验[J]. 中国科技信息，2016(21)：90-92.

[4] 薛裕颖，段希义，潘玉民. 时序与回归预测方法比较研究[J]. 黑龙江科技信息，2015(19)：129-131.

[5] 李 为，李一平，封锡盛. 基于卡尔曼滤波预测的无偏量测转换方法[J]. 控制与决策，2015，30(2)：229-234.

[6] 翟 静，曹 俊. 基于时间序列 ARIMA 与 BP 神经网络的组合预测模型[J]. 统计与决策，2016(4)：29-32.

[7] 王建平，胡 益，侍洪波. 基于高阶偏最小二乘的间歇过程建模[J]. 化工学报，2014，65(9)：3527-3534.

[8] ZHAO Qibin，CAIAFA C F，MANDIC D P，et al. Higher order partial least squares (HOPLS)：a generalized multilinear

（下转第 123 页）