Выбор тензорных представлений для прогнозирования по мультимодальным измерениям

Алсаханова Н. Ю., Стрижов В. В.

alsahanova.nyu@phystech.edu, strijov@phystech.edu

В данной работе решается задача прогнозирования по измерениям с разнородных источников путем выбора оптимального тензорного представления исходных данных. Оптимальное представление предлагается получать согласованием мультимодальных измерений. Далее строится прогностическая модель путем согласования независимых и целевой переменных.

Ключевые слова: *декодирование сигналов, электрокортикограмма, метод частичных наименьших квадратов, тензорый метод главных компонент. .*

1 Введение

Исходные данные в области декодирования сигналов [1], [2], [3] являются многомерными и сильно избыточными. В следствие чего модели, построенные на таких данных, неустойчивы. Так же значительно увеличивается время вычислений. Для решения этих проблем применяют модели отбора признаков [4], [5] и снижения размерности [6].

Часто в задаче декодирования имеется несколько входных сигналов, мультимодальные измерения. Например, в задаче восстановления траектории движения руки имеется 64 входных сигнала электрокортикограмм [2]. Для решения задачи с мультимодальными измерениями применяются две различные группы подходов, описанные в [7]. Одна группа состоит из методов, переводящих мультимодальные измерения в одно низкоразмерное представление. Другая группа методов путем согласования мультимодальных измерений проектирует их в низкоразмерное пространство, сохраняя их количество. Например, применяется канонический корреляционный анализ (CCA) и некоторые его модификации [8], [9], [10]. Так же согласование мультимодальных измерений проводится путем применения метода частичных наименьших квадратов (PLS) и его разновидностей [11], [12].

При прогнозировании временных рядов исходные данные имеют высокую размерность, так же в пространствах целевой и независимой переменных есть скрытые зависимости. Черезмерно высокая размерность пространств и наблюдаемая множественная корреляция приводит к неустойчивости прогностичекой модели. В данном случае регрессионная модель, основанная на методе частичных наименьших квадратов широко используется для восстановления зависимостей между наборами данных [13], [14], [15], [16]. Для построения устойчивой прогностической модели так же используют метод отбора признаков, основанный на задаче квадратичного программирования (QPFS) [3]. Данный метод максимизирует релевантность и минимизирует избыточность признаков. В статье [5] показано, что метод QPFS превосходит многие существующие методы отбора признаков для задачи одномерной регрессии.

Все вышеуказанные методы используют матричное представление данных для прогнозирования временных рядов. Но в задаче с мультимодальными измерениями данные представляются в тензорном виде. Можно решать задачу путем перехода от тензоров к матрицам, но это снижает интерпретируемость и увеличивает избыточность данных. Поэтому для работы с тензорными данными предлагались различные модели [17], [18]. Были предложены тензорные модификации метода главных компонент (tensor KPCA и MPCA) [19], [20], канонического корреляционного анализа (tensor CCA и deep tensor CCA) [21], [9] и метода частичных наименьших квадратов (HOPLS) [22]. В статьях [23], [24] были рассмотрены нелинейные версии тензорного метода частичных наименьших квадратов.

Когда мы имеем избыточные высокоразмерные данные, применение методов уменьшения размерности без согласования, таких как PCA, QPFS или MPCA, перед согласованием исходных данных с целевой переменной позволяет улучшить качество прогнозирования. Например, в статье [25] было показано, что при классификации изображений применение MPCA перед линейным дискриминантным анализом (LDA) позволило с большей точностью классифицировать изображения. В данной работе предлагается сначала находить оптимальные тензорные представления исходных данных с помощью тензорного метода главных компонент, а затем проводить согласование с целевой переменной с помощью PLS и его модификаций. Это позволить ускорить согласование и увеличит точность модели.

2 Постановка задачи

2.1 Мультимодальная регрессия

Пусть выборка имеет вид $(\underline{\mathbf{X}},\mathbf{Y})$:

$$\underline{\mathbf{X}} \in \mathbb{R}^{M\times n_1\times\ldots\times n_D}, \qquad \mathbf{Y} \in \mathbb{R}^{M\times K}, \tag{1}$$

где $y_{:k} = y_{:k}(t),\ k = 1,\ldots,K$ - временные ряды.

Тензор признаков $\underline{\mathbf{X}}$ имеет $D+1$ размерность. Чтобы восстановить целевые временные ряды, можно матрицизовать тензор признаков $\underline{\mathbf{X}}$ по первой размерности:

$$\underline{\mathbf{X}}_{(1)} = \left[vec\left(\underline{\mathbf{X}}_1\right)^\top,\ldots,vec\left(\underline{\mathbf{X}}_M\right)^\top\right]^\top \in \mathbb{R}^{M\times(n_1\cdot\ldots\cdot n_D)} \tag{2}$$

таким образом задача сводится к задаче мультимодальной регрессии, где $\mathbf{X} \in \mathbb{R}^{M\times(n_1\cdot\ldots\cdot n_D)}$ - исходная матрица, полученная матрицизацией исходного тензора, $\mathbf{Y} \in \mathbb{R}^{M\times K}$ - целевая матрица. Требуется построить прогностическую модель $\mathbf{f}: \mathbb{X} \to \mathbb{Y}$. Данная модель должна быть оптимальной $\mathbf{f}^*$, то есть минимизирующей некоторую функцию ошибки $\mathcal{L}$:

$$\mathbf{f}^* = \arg\min_{\mathbf{f}} \mathcal{L}\left(\mathbf{f},\ \mathbf{X},\ \mathbf{Y}\right) \tag{3}$$

Будем рассматривать только класс параметрических моделей $\mathbf{f}(\mathbf{X}, \boldsymbol{\Theta})$, где $\boldsymbol{\Theta}$ - матрица параметров, для сужения пространства поиска моделей. Тогда задача (3) сводится к задаче поиска оптимальных параметров:

$$\Theta^* = \arg\min_{\boldsymbol{\Theta}} \mathcal{L}(\Theta,\ \mathbf{X},\ \mathbf{Y}) \tag{4}$$

В даной работе рассматриваемые пространства исходной и целевой переменных избыточны. В таком случае решение задачи (4) неустойчиво. Например, при восстановлении линейной регрессии, когда $\mathbf{f}(\mathbf{X}, \Theta)$ линейная:

$$\mathbf{Y} = \mathbf{X}\Theta + \varepsilon, \qquad \Theta \in \mathbb{R}^{(n_1 \cdot \ldots \cdot n_D) \times K} \tag{5}$$

Оптимальные параметры Θ^* определяются минимизацией функции ошибки $\mathcal{L}(\Theta, \mathbf{X}, \mathbf{Y})$. Рассмотрим квадратичную фукцию потерь:

$$\mathcal{L}(\Theta, \mathbf{X}, \mathbf{Y}) = \left\| \mathbf{Y} - \mathbf{X}\Theta \right\|_2^2 \to \min_{\theta} \tag{6}$$

Решением задачи оптимизации (6) является:

$$\Theta = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{Y} \tag{7}$$

Линейная зависимость между признаками приводит к неустойчивости решения задачи (6), так как матрица $\mathbf{X}^\top \mathbf{X}$ близка к сингулярной. Решением данной проблемы являются методы снижения размерности, описываемые в пункте 1. Они позволяют избежать сильной линейной зависимости между признаками.

Так же если в выборке признаки представляют собой не матрицу $\mathbf{X} \in \mathbb{R}^{M \times (n_1 \cdot \ldots \cdot n_D)}$, а $D+1$-мерный тензор $\underline{\mathbf{X}}$, то процесс матрицизации приводит к потери информации, заложенной в тензорной структуре. Поэтому в данной работе так же исследуются тензорная регрессия 1 и тензорные методы снижения размерности пространства 1, 1, 1.

2.2 Тензорная регрессия

Тензорные представления данных часто очень полезны для смягчения проблемы небольшого размера выборки при построении прогностической модели. Это происходит за счет того, что переход к векторам и матрицам приводит чрезмерной подгонке для данных большого размера. Так же теряется информации для структурированных данных при матрицизации или векторизации данных. И так как тензоры это обобщение матриц и векторов, то тензорная регрессия может быть определена подобным образом, как (5):

$$\mathbf{y}_m = \langle \underline{\mathbf{X}}_m | \underline{\mathbf{W}} \rangle + \varepsilon \tag{8}$$

где $\langle \underline{\mathbf{X}}_m | \underline{\mathbf{W}} \rangle$ - тензорная свертка по первым D размерностям D-мерного исходного тензора $\underline{\mathbf{X}}_m \in \mathbb{R}^{n_1 \times \ldots \times n_D}$ и $(D+1)$-мерного тензора параметров модели $\underline{\mathbf{W}} \in \mathbb{R}^{n_1 \times \ldots \times n_D \times K}$, $\varepsilon \in \mathbb{R}^K$ - ошибка, $\mathbf{y}_m \in \mathbb{R}^K$ - целевой вектор.

k-ый элемент вектора, получившегося после тензорной свертки $\underline{\mathbf{X}}_m$ и $\underline{\mathbf{W}}$ считается следующим образом:

$$\langle \underline{\mathbf{X}}_m | \underline{\mathbf{W}} \rangle_k = \sum_{i_1=1}^{n_1} \cdots \sum_{i_D=1}^{n_D} x_{i_1,\ldots,i_D} w_{i_1,\ldots,i_D,k} \tag{9}$$

Оптимальный тензор параметров $\underline{\mathbf{W}}^*$ находится путем минимизации квадратичной функции ошибки:

$$\underline{\mathbf{W}}^* = \arg\min_{\underline{\mathbf{W}}} \sum_{m=1}^{M} \left\| \mathbf{y}_m - \langle \underline{\mathbf{X}}_m | \underline{\mathbf{W}} \rangle \right\|_2^2 \tag{10}$$

Тензор параметров $\underline{\mathbf{W}}$ имеет много размерностей, поэтому чтобы решить задачу оптимизации 10 и сделать это наиболее эффективным способом предлагается представить тензор параметров в раложении Такера:

$$\underline{\mathbf{W}} \approx \underline{\mathbf{G}} \times_1 \mathbf{U}^{(1)} \ldots \times_D \mathbf{U}^{(D)} \times_{D+1} \mathbf{U}^{(D+1)} \tag{11}$$

где $\underline{\mathbf{G}}$ - центральный тензор меньшего размера, $\mathbf{U}^{(i)}$ -

2.3 Уменьшение размерности

2.4 Предлагаемая модель

$$\mathcal{L}\left(\mathbf{f_1}, \mathbf{f_2}, \underline{\mathbf{X}}, \mathbf{Y}\right) \to \min_{\mathbf{f_1}, \mathbf{f_2}} \tag{12}$$

3 Метод главных компонент

3.1 Мультилинейный метод главных компонент

3.2 Метод главных компонент для трехмерного случая

4 Метод частичных наименьших квадратов

4.1 Тензорный метод частичных наименьших квадратов

5 Эксперимент

5.1 Метрики

5.2 Данные

5.3 Результаты

6 Заключение

Литература

[1] Jean Faber Alexander Wyss Napoleon Torres Corinne Mestais Alim Louis Benabid Andrey Eliseyev, Cecile Moro and Tetiana Aksenova. L1-penalized n-way pls for subset of electrodes selection in bci experiments. *Journal of neural engineering*, 9(4), 2012.

Таблица 1 Данные

Название	Источники	Предсказываемый сигнал
Synthetic		
ECoG	$\underline{\mathbf{X}} \in \mathbb{R}^{t\times 32\times 27}$ - 32 временных ряда с датчиков, 27 частот	$\mathbf{Y} \in \mathbb{R}^{t\times 3}$ - 3 координаты положения руки
My data	$\underline{\mathbf{X}} \in \mathbb{R}^{t\times 75\times 2}$	$\mathbf{Y} \in \mathbb{R}^{t\times 75}$
BerkeleyMHAD	$\underline{\mathbf{X}} \in \mathbb{R}^{t\times 3\times 43}$	$\underline{\mathbf{Y}} \in \mathbb{R}^{t\times 3\times 6}$
BerkeleyMHAD-img	$\underline{\mathbf{X}} \in \mathbb{R}^{t\times 15\times 35\times 4}$	$\underline{\mathbf{Y}} \in \mathbb{R}^{t\times 3\times 6}$

Таблица 2 Результаты экспериментов

Данные	Модель	Количество компонент	RMSE	Время, сек
ECoG	PLS	14	0.9578	22
	QPFS+PLS	13	0.9526	116
	HOPLS	27	0.9564	39
	MPCA+PLS	**2**	**0.9525**	**3**
	MPCA+HOPLS	29	0.9564	43
My data	PLS	**5**	0.8224	**0.2**
	QPFS+PLS	**5**	0.8998	7.7
	HOPLS	45	0.6275	1.9
	MPCA+PLS	6	0.7814	0.3
	MPCA+HOPLS	25	**0.5956**	3

[2] Andrey Eliseyev and Tatiana Aksenova. Stable and artifact-resistant decoding of 3d hand trajectories from ecog signals using the generalized additive model. *Journal of neural engineering*, 11(6), 2014.

[3] Anastasia Motrenko and Vadim Strijov. Multi-way feature selection for ecog-based brain-computer interface. *Expert Systems with Applications*, 114:402–413, 2018.

[4] Lars Snipen Tahir Mehmood, Kristian Hovde Liland and Solve Saebo. A review of variable selection methods in partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 118(12):62–69, 2012.

[5] Alexandr Katrutsa and Vadim Strijov. Comprehensive study of feature selection methods to solve multicollinearity problem according to evaluation criteria. *Expert Systems with Applications*, 76:1–11, 2017.

[6] Alexandr Katrutsa and Vadim Strijov. Stress test procedure for feature selection algorithms. *Chemometrics and Intelligent Laboratory Systems*, 142:172–183, 2015.

[7] Yingming Li, Ming Yang, and Zhongfei Zhang. A survey of multi-view representation learning. *IEEE transactions on knowledge and data engineering*, 31(10):1863–1883, 2018.

[8] Adrian Benton, Huda Khayrallah, Biman Gujral, Drew Reisinger, Sheng Zhang, and Raman Arora. Deep generalized canonical correlation analysis. *CoRR*, abs/1702.02519, 2017.

[9] Hok Shing Wong, Li Wang, Raymond H. Chan, and Tieyong Zeng. Deep tensor cca for multi-view learning. *CoRR*, abs/2005.11914, 2020.

[10] Heather D. Couture, Roland Kwitt, J. S. Marron, Melissa A. Troester, Charles M. Perou, and Marc Niethammer. Deep multi-view learning via task-optimal cca. *CoRR*, abs/1907.07739, 2019.

[11] Andrey Eliseyev and Tetiana Aksenova. Penalized multi-way partial least squares for smooth trajectory decoding from electrocorticographic (ecog) recording. *PloS one*, 11(5), 2016.

[12] Roman Rosipal. Nonlinear partial least squares an overview. *Chemoinformatics and advanced machine learning perspectives: complex computational methods and collaborative techniques*, pages 169–189, 2011.

[13] Roman Rosipal and Nicole Kramer. Overview and recent advances in partial least squares. *C. Saunders et al. (Eds.) SLSFS 2005, LNCS 3940*, page 34–51, 2006.

[14] Alexander E. Stott, Bruno Scalzo Dees, Ilia Kisil, and Danilo P. Mandic. A class of multidimensional nipals algorithms for quaternion and tensor partial least squares regression. *Signal Process.*, 160:316–327, 2019.

[15] Solve Sæbø Tahir Mehmood and Kristian Hovde Liland. Comparison of variable selection methods in partial least squares regression. *Journal of Chemometrics*, 2020.

[16] Petre Manolescu Julien Lauzon-Gauthier and Carl Duchesne. The sequential multi-block pls algorithm (smb-pls): Comparison of performance and interpretability. *Chemometrics and Intelligent Laboratory Systems*, (180):72–83, 2018.

[17] Eric F. Lock. Tensor-on-tensor regression. *Journal of Computational and Graphical Statistics*, 27(3):638–647, 2018.

[18] Kamran Paynabar Mostafa Reisi Gahrooei, Hao Yan and Jianjun Shi. Multiple tensor-on-tensor regression: An approach for modeling processes with heterogeneous sources of data. *Technometrics*, 2020.

[19] Johan A. K. Suykens Lynn Houthuys. Tensor learning in multi-view kernel pca. *Artificial Neural Networks and Machine Learning - ICANN 2018*, 11140, 2018.

[20] Haiping Lu, Konstantinos N. Plataniotis, and Anastasios N. Venetsanopoulos. Multilinear principal component analysis of tensor objects for recognition. In *ICPR (2)*, pages 776–779. IEEE Computer Society, 2006.

[21] Yong Luo, Dacheng Tao, Yonggang Wen, Kotagiri Ramamohanarao, and Chao Xu. Tensor canonical correlation analysis for multi-view dimension reduction. *CoRR*, abs/1502.02330, 2015.

[22] Qibin Zhao, Cesar F Caiafa, Danilo P Mandic, Zenas C Chao, Yasuo Nagasaka, Naotaka Fujii, Liqing Zhang, and Andrzej Cichocki. Higher order partial least squares (hopls): a generalized

multilinear regression method. *IEEE transactions on pattern analysis and machine intelligence*, 35(7):1660–1673, 2012.

[23] Y. Xu X. Feng and Y. Meng. Short-term load forecasting with tensor partial least squares-neural network. *Energies*, 990(12), 2019.

[24] Qibin Zhao, Liqing Zhang, and Andrzej Cichocki. Multilinear and nonlinear generalizations of partial least squares: an overview of recent advances. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, 4(2):104–115, 2014.

[25] Jun Huang, Kehua Su, Jamal El-Den, Tao Hu, and Junlong Li. An mpca/lda based dimensionality reduction algorithm for face recognition. *Mathematical Problems in Engineering*, 2014:1–12, 08 2014.