

教師あり主成分分析を用いた動的モード分解

Dynamic mode decomposition using supervised principal component analysis

尾藤 岳仁^{*1}

Bito Takehito

河原 吉伸^{*1*2}

Kawahara Yoshinobu

鷺尾 隆^{*1}

Washio Takashi

^{*1}大阪大学 産業科学研究所

The Institute of Scientific and Industrial Research, Osaka University

^{*2}理化学研究所 革新知能統合研究センター

Center for Advanced Intelligence Project, RIKEN

Dynamic mode decomposition(DMD) is a data-driven method for representing high-dimensional, nonlinear dynamical systems. DMD extracts key low-rank spatiotemporal features of the high-dimensional systems. However, since DMD is an unsupervised method and, thus, cannot incorporate label information into it even when such information is available. In this paper, we propose a framework to incorporate supervised information into DMD analyses. Experimental results show the effectiveness of performing classification tasks using modes obtained by the proposed method.

1. はじめに

何らかの基準に基づいてデータを複数のモードに分解することは、複雑なシステムを解析するために用いられる標準的な手法の一つである。動的モード分解 (Dynamic Mode Decomposition, DMD) は、多次元時系列データを周波数と増減率の情報をもった複数のモードに分解する。このアルゴリズムは時間および空間上の次元削減手法としても知られ、高次元動的システムの重要な低ランクの時間と空間上の特徴を抽出する。それにより、データの空間上の構造およびそれに関連する時間軸上の物理的な解釈を可能にする。DMD に対応する確率モデルを開発しベイズ的に拡張した手法 [1] やモードをスパースにし、ダイナミクスに関わるモードを選択する手法 [2] など DMD は、幅広く応用され注目を集めている。

一方で、DMD は教師なし学習である。したがって、ラベル情報があるようなデータの場合でも、その情報を有効にできないため、ラベル情報に関連した特に興味のあるモードを導くことができない可能性がある。一般の DMD アルゴリズムでは特異値分解により分散最大方向へデータを射影し、ダイナミクスを表す行列の固有値分解を行いモード分解をしていた。この特異値分解による射影では、目的変数については何も考慮していない。しかし、教師情報がある場合、目的変数との依存関係を考慮した方向へデータを射影しモード分解をすることが好ましい。分類問題を想定すると、分類に最も影響を及ぼすモードを用いて分類問題を解くことで精度の向上が期待できる。このような目的のモードは分散最大方向の射影によって導かれるとは限らない。

そこで本研究では、説明変数と目的変数の依存性を最大とする空間へデータを射影する教師あり PCA [3] を取り入れることで、目的変数と最も関連したモードを導くことができるように改良した。そして教師情報を考慮した DMD によって導かれたモードによるデータの類似性を可視化することにより、提案手法の有効性を示す。

本稿は以下のように構成されている。2. 節では、DMD とそのアルゴリズムについて示す。3. 節では、教師情報を取り入れた DMD の提案手法を述べ、3.1 節において説明変数と目的変数の依存性を最大とする射影を求める教師あり主成分分析 (教師あり PCA) を示す。3.2 節では、教師あり PCA によって求めた射影を用いた DMD アルゴリズムを示す。4. 節で提案手法の有用性を時系列データの類似性の可視化によって示し、5. 節で本稿の内容をまとめ、結論を述べる。

2. 動的モード分解

DMD では動的システムから得られる多次元時系列データを時間軸上では周波数と増減率を持つ波に、空間上ではその波に対する各属性の寄与率を持つ複数のモードに分解する。時系列データ

$$X = \begin{bmatrix} | & | & \cdots & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_m \\ | & | & & | \end{bmatrix} \quad (1)$$

から、次のような 2 つのデータ行列を生成する。

$$X_1 = \begin{bmatrix} | & | & \cdots & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_{m-1} \\ | & | & & | \end{bmatrix} \quad (2)$$

$$X_2 = \begin{bmatrix} | & | & \cdots & | \\ \mathbf{x}_2 & \mathbf{x}_3 & \cdots & \mathbf{x}_m \\ | & | & & | \end{bmatrix} \quad (3)$$

ここで $\mathbf{x}_k \in \mathbb{R}^p$ は時刻 k における動的システムから得られる観測量であり、 p は観測点 (属性) の数を表す。 X_2 は X_1 に対して 1 ステップ後の状態を表しており、DMD では $X_2 \approx AX_1$ つまり、

$$A = X_2 X_1^\dagger \quad (4)$$

連絡先: 大阪大学産業科学研究所 知能推論研究分野

〒 567-0047 大阪府茨木市美穂ヶ丘 8-1

E-mail: bito@ar.sanken.osaka-u.ac.jp

として線形作用素の有限データによる近似 A の固有値分解を求める。ただし † は Moore-Penrose の疑似逆行列を表す。 A の固有値が時間軸上の周波数および増減率を示し、固有ベクトルが特定の固有値に対する空間上の寄与率を示している。その固有ベクトルは DMD モードと呼ばれており、各波に対する各属性の寄与率を示しているため空間上の構造を表している。DMD アルゴリズムを Algorithm1 に示す [4]。

Algorithm 1 DMD アルゴリズム

1: 特異値分解

$$X_1 \approx U \Sigma V^* \quad (5)$$

ただし、 $*$ は複素共役転置を表し、 $U \in \mathbb{C}^{p \times r}$ 、 $\Sigma \in \mathbb{C}^{r \times r}$ 、 $V \in \mathbb{C}^{m-1 \times r}$ 。 r は X_1 に対する特異値分解によって次元削減された次元である。 U と V の列は直交であるため、 $U^*U = I, V^*V = I$ となる。

2: $\tilde{A}$ を求める。

$$\tilde{A} = U^*AU = U^*X_2V\Sigma^{-1} \quad (6)$$

3: $\tilde{A}$ の固有値分解

$$\tilde{A}W = W\Lambda \quad (7)$$

ここで、 W の列は $\tilde{A}$ の固有ベクトルであり、 Λ は対角成分に対応する固有値を持つ対角行列である。

4: DMD モードの計算

A の固有値は Λ に対応し、DMD モードは以下で与えられる [4]。

$$\Phi = X_2V\Sigma^{-1}W\Lambda^{-1} \quad (8)$$

3. 提案手法

DMD の第二ステップ (6) 式では、特異値分解によりノイズの影響を抑える空間にデータを射影していた。しかし、クラス分類問題のように教師情報のある問題では、分類に適したモードを導くことで分類精度の向上が期待できる。したがって、本稿では DMD の第二ステップにおいて、特異値分解によって求まる U による射影の代わりに、教師情報を取り入れ説明変数と目的変数の依存性が最大となる空間に射影する教師あり PCA によってモードを導くことを提案する。また、DMD モードはモード分解したときの各波に対するそれぞれの属性の寄与率を表しており、空間上のコヒーレントを示している。つまり、DMD モードにより分類を行うことで、動的システムのダイナミクスを反映した分類を考えることができる。3.1 節で教師あり PCA の概要を説明し、3.2 節で教師あり PCA を用いて教師情報を考慮した DMD アルゴリズムを示す。

3.1 教師あり PCA

教師あり PCA [3] では、説明変数を目的変数との依存性が最大となる空間へ射影する U_s を求める。説明変数 $\mathcal{X}$ と目的変数 $\mathcal{Y}$ の依存度を測るために Hilbert-Schmidt independence criterion (HSIC) [5] を導入した。HSIC では $\mathcal{X}, \mathcal{Y}$ のそれぞれの再生核ヒルベルト空間 $\mathcal{F}, \mathcal{G}$ に関連した相互共分散オペレータの Hilbert-Schmidt ノルムを計算し依存度を測っている。観測データ $\mathcal{Z} := \{(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_N, \mathbf{y}_N)\} \subseteq \mathcal{X} \times \mathcal{Y}$ について、

HSIC の推定値は

$$HSIC(\mathcal{Z}, \mathcal{F}, \mathcal{G}) = (N-1)^2 \text{tr}(HKHL) \quad (9)$$

で与えられる。ただし、 $H, K, L \in \mathbb{R}^{N \times N}$ 、 $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ 、 $L_{ij} = l(\mathbf{y}_i, \mathbf{y}_j)$ 、 $H_{ij} = I - N^{-1}\mathbf{e}\mathbf{e}^\top$ であり、 k, l はそれぞれ $\mathcal{F}, \mathcal{G}$ のカーネル関数である ($\mathbf{e}$ はすべての要素が 1 である N 次元ベクトル)。

観測データ行列 $X = [\mathbf{x}_1 \dots \mathbf{x}_N]$ 、 $Y = [\mathbf{y}_1 \dots \mathbf{y}_N]$ とする。教師あり PCA は射影されたデータ $U^\top X$ と結果 Y の依存性が高い部分空間を求めるため HSIC、 $\text{tr}(KHLH)$ の最大化をする。ここで、 K は $U^\top X$ のカーネル (e.g. $X^\top U U^\top X$) であり、 L は Y のカーネル (e.g. $Y^\top Y$) である。クラス分類問題を考える場合は Y のカーネルとして、デルタカーネル $l(\mathbf{y}_i, \mathbf{y}_j) = \delta(\mathbf{y}_i, \mathbf{y}_j)$ を用いる。以上より、目的関数は

$$\begin{aligned} \text{tr}(HKHL) &= \text{tr}(HX^\top U U^\top XHL) \\ &= \text{tr}(U^\top XHLHX^\top U) \end{aligned} \quad (10)$$

と変形できる。ここで各データは属性が相関していない空間に射影されるため、 U は直交射影とする。したがって、最適化問題は以下となる。

$$\begin{aligned} \arg \max_U \quad & \text{tr}(U^\top XHLHX^\top U) \\ \text{subject to} \quad & U^\top U = I \end{aligned} \quad (11)$$

対称行列 $Q = XHLHX^\top$ としたときの固有値 $\lambda_1 \leq \dots \leq \lambda_p$ 、対応する固有ベクトルを $\mathbf{v}_1, \dots, \mathbf{v}_p$ とすると、制約を満たす目的関数の最大値は $\lambda_p + \lambda_{p-1} + \dots + \lambda_{p-d+1}$ であり、最適解は $U = [\mathbf{v}_p \mathbf{v}_{p-1} \dots \mathbf{v}_{p-d+1}]$ である [6]。ただし、 d は出力空間の次元である。

教師あり PCA のアルゴリズムを Algorithm 2 に示す。

Algorithm 2 教師あり PCA アルゴリズム

Input: X, Y, L, d

Output: U

- 1: $H \leftarrow I - n^{-1}\mathbf{e}\mathbf{e}^\top$
 - 2: $Q \leftarrow XHLHX^\top$
 - 3: $U \leftarrow Q$ の上位 d 個の固有値に対応する固有ベクトル
-

3.2 DMD への適用

本稿では、ラベル付き多次元時系列データへの DMD の適用を考えるため、 N 個の時系列 $\{X^{(i)}, y^{(i)}\}_{i=1}^N$ について、各時系列 $X^{(i)}$ は (1) で示すような p 次元時系列を表す行列であり、 $y^{(i)}$ はその時系列のクラスとする。提案手法では 2 つのステップにより教師情報を反映した DMD モードを求める。第一ステップは 3.1 節で示した教師あり PCA によりデータを教師情報を反映した空間に射影する U_s を求める。ここで時系列データに対し教師あり PCA を実行するために、時系列は定常性を仮定する。つまり、時刻によってダイナミクスは変化せず、どの時刻においても各属性とラベルの依存関係は変化しないと仮定する。そして、全時刻に共通な U_s を求める。第二ステップでは、各時系列を共通な U_s により射影し、2. 節で示した DMD により DMD モードを求める。

まず第一ステップでは、教師あり PCA のために以下のようなデータ行列 X_s を生成する。

$$X_s := [X_1^{(1)} \ X_1^{(2)} \ \dots \ X_1^{(N)}] \quad (12)$$

$X_1^{(i)}$ は $X^{(i)}$ に関して (2) のような時刻 1 から $m-1$ の時系列を表す行列である。これを用いて最適化問題 (11) は以下となる。

$$\begin{aligned} \arg \max_{U_s} \quad & \text{tr}(U_s^\top X_s H L_s H X_s^\top U_s) \\ \text{subject to} \quad & U_s^\top U_s = I \end{aligned} \quad (13)$$

ただし, X_s は時系列データとして拡張してあるため, L_s は以下のように定義する。

$$L_s = \begin{bmatrix} l_{11} & \dots & l_{1N} \\ \vdots & \ddots & \vdots \\ l_{N1} & \dots & l_{NN} \end{bmatrix}. \quad (14)$$

l_{ij} は $y_i = y_j$ のとき $1^{(m-1 \times m-1)}$ (すべての要素が 1 の $m-1$ 行 $m-1$ 列の行列), $y_i \neq y_j$ のとき $0^{(m-1 \times m-1)}$ とする。以上の最適化問題から最適化問題 (11) と同様に固有値分解により U_s を求める。

第二ステップでは各観測データに対して DMD を実行する。Algorithm1 に則り教師ありに拡張した手続きを示していく。ここで, 提案手法では教師情報を考慮するため, 共通な U_s を用いてデータを射影し DMD の計算を行っている。各時系列の特異値分解で得られる $U^{(i)}$ と教師情報を考慮した空間へ射影する U_s が異なることに注意されたい。

1. 特異値分解: $X_1^{(i)} \approx U^{(i)} \Sigma^{(i)} V^{(i)*}$
2. 教師情報を取り入れた空間へ射影:
 $\tilde{A}_s^{(i)} = U_s^\top A^{(i)} U_s = U_s^\top X_2^{(i)} V^{(i)} \Sigma^{(i)} U^{(i)*} U_s$
3. $\tilde{A}^{(i)}$ の固有値分解: $\tilde{A}^{(i)} W^{(i)} = W^{(i)} \Lambda^{(i)}$
4. DMD モード $\Phi^{(i)}$:
 $\Phi^{(i)} = A^{(i)} U_s W^{(i)} \Lambda^{(i)-1}$
 $= X_2^{(i)} V^{(i)} \Sigma^{(i)} U^{(i)*} U_s W^{(i)} \Lambda^{(i)-1}$

また, 各観測データにおいて教師あり PCA の射影による $\mathbf{x}_{k+1}$ と $\mathbf{x}_k$ の関係についても議論する。時刻 k における観測データ $\mathbf{x}_k$ を教師情報を考慮した U_s により射影し $\tilde{A}_s$ をかけると,

$$\begin{aligned} \tilde{A}_s U_s^\top \mathbf{x}_k &= \tilde{A}_s \tilde{\mathbf{x}}_k \\ &= U_s^\top A U_s \tilde{\mathbf{x}}_k \\ &= U_s^\top X_2 V \Sigma^{-1} U^\top U_s (U_s^\top \mathbf{x}_k) \\ &= U_s^\top X_2 V \Sigma^{-1} U^\top \mathbf{x}_k \\ &= U_s^\top \mathbf{x}_{k+1} \end{aligned} \quad (15)$$

となり, $\tilde{A}_s$ が U_s によって次元削減された空間におけるダイナミクスを表していることが確かめられる。

以上をまとめて提案手法のアルゴリズムを Algorithm3 に示す。

4. 実験

教師情報を取り入れ, DMD モードによるクラス分類が効果的であることを示すために多次元尺度構成法 (MDS) によりデータの類似性の可視化を行った。使用したデータは CMU Graphics Lab Capture Database[7] の Subject 64(golf) である。DMD モードの順序に影響を受けず類似性を測るため

Algorithm 3 提案手法

Input: $\{X^{(i)}, y^{(i)}\}_{i=1}^N, L_s, d, r$

Output: $\{\Phi^{(i)}\}_{i=1}^N$

- 1: $H \leftarrow I - n^{-1} \mathbf{e} \mathbf{e}^\top$
- 2: $X_s \leftarrow [X_1^{(1)} X_1^{(2)} \dots X_1^{(N)}]$
- 3: $Q_s \leftarrow X_s H L_s H X_s^\top$
- 4: $U_s \leftarrow Q_s$ の上位 d 個の固有値に対応する固有ベクトル
- 5: **for** $i = 1$ to N **do**
- 6: $U^{(i)}, \Sigma^{(i)}, V^{(i)} \leftarrow X_1^{(i)}$ の特異値分解 ($X_1^{(i)} \approx U^{(i)} \Sigma^{(i)} V^{(i)*}$)
- 7: $\tilde{A}_s^{(i)} \leftarrow U_s^\top X_2^{(i)} V^{(i)} \Sigma^{(i)} U^{(i)*} U_s$
- 8: $W^{(i)} \leftarrow \tilde{A}_s^{(i)}$ の固有値分解 ($\tilde{A}_s^{(i)} W^{(i)} = W^{(i)} \Lambda^{(i)}$)
- 9: $\Phi^{(i)} \leftarrow X_2^{(i)} V^{(i)} \Sigma^{(i)} U^{(i)*} U_s W^{(i)} \Lambda^{(i)-1}$
- 10: **end for**

に, DMD モードの線型結合によって定義される空間の距離を kernel principal angle[8] を用いて定義した。DMD モード $\Phi^{(i)}, \Phi^{(j)}$ の QR 分解を $\Phi^{(i)} = Q_i R_i, \Phi^{(j)} = Q_j R_j$ とすると, $\Phi^{(i)}$ と $\Phi^{(j)}$ の principal angle の $\cos$ は $Q_i^\top Q_j$ の特異値 $\delta_1, \dots, \delta_k$ によって与えられる。したがって, 部分空間 Q_i を点と見なすことができるグラスマン多様体上のグラスマンベクトルを $\Psi(Q_i)$ として, Q_i, Q_j 間の距離 D_{ij} を以下のように定義した。

$$\begin{aligned} D_{ij} &= \|\Psi(Q_i) - \Psi(Q_j)\|^2 \\ &= \|\Psi(Q_i)\|^2 + \|\Psi(Q_j)\|^2 - 2\Psi(Q_i)^\top \Psi(Q_j) \\ &= 2(1 - \prod_{i=1}^k \sigma_i) \end{aligned} \quad (16)$$

上式の距離行列 D を入力として MDS を実行した際の DMD のみの出力を図 1, 教師情報を取り入れた提案手法の出力を図 2 に示す。

DMD では'swing' データはまとまっているが, 'placing tee', 'placing ball', 'placing up ball' のモーションは分離できていないとは言えない。しかし, 提案手法による可視化では, DMD で分離できなかった 3 つのモーションが分けられている。また, DMD に見られた'putt'の外れ値も修正されており, 教師情報を取り入れ DMD による分類が効果的であることが示せた。

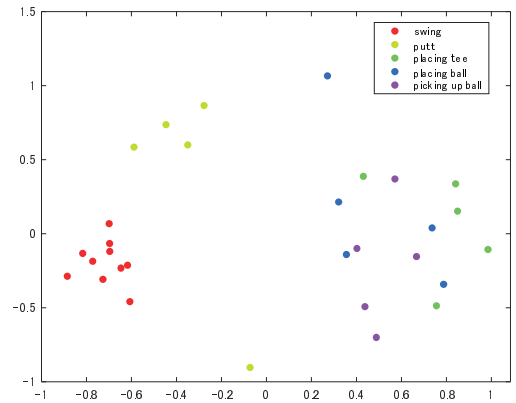


図 1: DMD による類似性の可視化

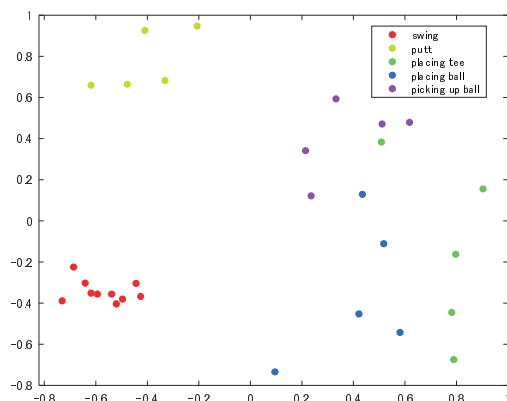


図 2: 提案手法による類似性の可視化

5. むすび

本稿では、DMD が教師なし学習であることから分類に適した DMD モードが抽出できない可能性があることに言及し、教師情報を取り入れるために教師あり PCA によってデータを射影し DMD を実行する手法の提案をした。また、実験を通して教師情報を用いることで DMD では分離できなかったデータをクラスごとに分離することができ、提案手法の有用性が示せた。

今後の課題は分類問題を解くことにより定量的な評価をすることやデータから直接類似性を測り分類する手法などその他の手法との比較が必要である。また、教師あり PCA[3] および DMD[9] はカーネル化された方法や、それに伴う非線形ダイナミクスの比較手法 [10] も提案されているため、提案手法もカーネル化することによりさらに高精度な結果が期待できる。

参考文献

- [1] N. Takeishi, Y. Kawahara, Y. Tabei, T. Yairi: Bayesian Dynamic Mode Decomposition, in *Proc. of the 26th Int'l Joint Conf. on Artificial Intelligence (IJ-CAI'17)*, pp. 2814–2821 (2017)
- [2] M. R. Jovanović, P. J. Schmid, J. W. Nicholas: Sparsity-promoting dynamic mode decomposition, *Physics of Fluids*, vol.26, 024103 (2014)
- [3] E. Barshas, A. Ghodsi, Z. Azimifar, M.Z. Jahromi: Supervised principal component analysis: Visualization, classification and regression on subspaces and submanifolds, *Pattern Recognition*, Vol. 44, pp. 1357–1371 (2011)
- [4] Jonathan H. Tu, Clarence W. Rowley, Dirk M. Luchtenburg, Steven L. Brunton, J. Nathan Kutz: On dynamic mode decomposition: theory and applications, *Journal of Computational Dynamics*, Vol. 1, No. 2, pp. 391–421 (2007)
- [5] A. Gretton, O. Bousquet, A.J. Smola, B. Schölkopf: Measuring statistical dependence with Hilbert-Schmidt norms, in *Proceedings Algorithmic Learning Theory(ALT)*, Vol. 3734, pp. 63–77 (2005)

- [6] H. Lutkepohl, Handbook of Matrices, John Wiley and Sons (1997)
- [7] CMU Graphics Lab Motion Capture Database: <http://mocap.cs.cmu.edu/>
- [8] L. Wolf, A. Shashua: Learning over sets using kernel principal angles, *Journal of Machine Learning Research*, Vol. 4, pp. 913–931 (2003)
- [9] Y. Kawahara: Dynamic Mode Decomposition with Reproducing Kernels for Koopman Spectral Analysis, *Advances in Neural Information Processing Systems*, vol. 29, pp. 911–919 (2016)
- [10] K. Fujii, Y. Inaba, Y. Kawahara: Koopman spectral kernels for comparing complex dynamics with application to multiagent in sports, in *Proc. of the 2017 European Conf. on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD'17)*, pp. 127–139 (2017)