特集記事

オートエンコーダによる低次元化と可視化*

尾亦 範泰**

Low-dimensionalization via Auto-encoder and Visualization

Noriyasu OMATA

1．はじめに

センシング技術の発達により，大量のデータが取得される時代になった．この時代において，人工知能と（情報）可視化はデータを人間が扱いやすい形に加工するという役割が共通しているように思われる[1),2)]．人工知能はデータから人間が「知りたい」情報を抽出しており，可視化はデータを人間が理解可能な「見やすい」形にして提供する．

人工知能，より端的には機械学習が進歩していく過程において，画像をはじめとする高次元のデータの扱いには苦難の歴史がある．入力となるデータの次元が大きいと，扱うモデルが複雑になりパラメータの数が膨大になってしまう．このようなモデルをやみくもに学習してしまうと，与えられたデータに特化することで新たなデータに対応できない過学習が起こる．このような現象は「次元の呪い」と呼ばれ，機械学習における悩みの種の一つであった[3),4)]．

この「呪い」を解決する方法として，高次元のデータを特徴量（feature），表現（representation），コード（code）などと呼ばれる低次元量に変換して用いる方法がある．低次元量として扱うことで，よりシンプルなモデルでパラメータ数を減らして扱うことができる．旧来どのような値を特徴量として用いるかは，人間がデータの特性を考慮して決定する必要があった．この作業は経験的な部分によるところが大きく，高度な人間の判断が必要であったため，機械学習における「匠の技」の側面があった．たとえば，画像認識の分野でよく用いられるSIFT特徴量は，回転やスケールの変化に対して不変な特徴量が望ましいという人間の判断に基づいて設計された[5)]．

これを自動化する一つの方策として示された手法が，オートエンコーダ（自己符号化器）[6)]である．オートエンコーダの思想はシンプルで，低次元の特徴量から元のデータを復元しようとしたとき，その復元の誤差を最小にできるような特徴量を用いるというものである．オートエンコーダによって抽出された特徴を用いることで，直接データを扱うよりも性能が上がることが様々な分野において報告された[7)]．

一方で，可視化においても「良い」特徴量を見つけて用いることは大切であるといえる．なぜなら人間の認識能力には限界があり，可視化して認識できる情報の量には限りがあるからである．そのため，多量のデータの中から可視化するべき特徴量を選択する必要がある．この選択によって可視化の結果は大きく変わる．しかしながら，可視化の分野において，特徴量を自動選択するといった試みは少ないように思われる．

本稿では，人工知能分野で用いられてきた，データから特徴量を自動抽出するための構造であるオートエンコーダについて述べる．まず，オートエンコーダについて概観し，いくつかの構造について説明する．さらに，非定常流れのデータに対してオートエンコーダを適用し，抽出された特徴量をもとに流れの時間変化を可視化した応用例について述べる．

2．オートエンコーダ概観

本節では，機械学習の分野において特徴抽出の手法として広く用いられてきたオートエンコーダについて，その概要を述べる．まずオートエンコーダについて定式化したのち，可視化分野における手法との関連性を指摘したうえで，その発展の経緯を概観する．

2.1 オートエンコーダの定式化

オートエンコーダは，データから特徴量を計算するための関数を取得する方策のひとつである．解析対象であるデータを $\mathbf{x}_t\,(t=1,\ldots,N)$ とする．それぞれのデータは M 次元のベクトルであるとして，$\mathbf{x}_t=(x_{t1},\ldots,x_{tM})$ と表記する．取得したい特徴量の次元を $H(<M)$ として，それぞれのデータに対する特徴量を $\mathbf{h}_t=(h_{t1},\ldots,h_{tH})$ とする．

オートエンコーダは，エンコーダ関数とデコーダ関数によって構成される．エンコーダ関数は，

$$\mathbf{h}=\mathrm{Encoder}(\mathbf{x}) \tag{1}$$

のように，データを特徴量に変換する関数である．一方，デコーダ関数は特徴量から元のデータを「復元」する関

* 原稿受付　2018年7月30日
** 正会員　東京大学 大学院工学系研究科
（〒113-8656　東京都文京区本郷7-3-1，
E-mail：omata@nakl.t.u-tokyo.ac.jp）

数である．復元されたデータを $\mathbf{y}_t=(y_{t1},\dots,y_{tM})$ とすると，

$$\mathbf{y}=\mathrm{Decoder}(\mathbf{h}) \tag{2}$$

のように表される．これらの関数はパラメータによって表現される関数であり，一般にはニューラルネットワークを用いる．エンコーダ関数とデコーダ関数のパラメータをまとめて $\boldsymbol{\theta}$ で表す．たとえば，それぞれの関数を

$$\begin{aligned}\mathrm{Encoder}(\mathbf{x})&=f(\mathbf{W}\mathbf{x}+\mathbf{b}),\\ \mathrm{Decoder}(\mathbf{h})&=f(\mathbf{W}^{\mathrm{T}}\mathbf{h}+\mathbf{c})\end{aligned} \tag{3}$$

としたとき，$\boldsymbol{\theta}=(\mathbf{W},\mathbf{b},\mathbf{c})$ である．

与えられたデータ $\mathbf{x}$ をエンコーダ関数に入れて得られる特徴量 $\mathbf{h}$ は，デコーダ関数に入れられることで $\mathbf{y}$ に「復元」される．しかしながら，この「復元」は一般に完全なものではなく誤差を含む．オートエンコーダは，この誤差を最小化するような変換が，良い特徴量を表すであろうという思想に基づく．

復元の「誤差」を測る誤差関数 $E(\mathbf{x},\mathbf{y})$ を考え，与えられたデータに対し特徴量から復元されたデータの誤差の総和 L を損失関数と呼ぶ．オートエンコーダは，この損失が最小になるように，エンコーダ関数とデコーダ関数のパラメータ $\boldsymbol{\theta}$ を，

$$\mathrm{minimize}_{\theta}\, L=\sum_{t=1}^{N} E(\mathbf{x}_t,\mathbf{y}_t) \tag{4}$$

のように最適化することで得られる．誤差関数には 2 乗誤差

$$E(\mathbf{x},\mathbf{y})=|\mathbf{x}-\mathbf{y}|^2=\sum_{m=1}^{M}|x_m-y_m|^2 \tag{5}$$

が広く用いられる．その他にも，データ $\mathbf{x}$ が確率やカテゴリを表しており 0 ～ 1 の値を取る場合には，クロスエントロピー

$$E(\mathbf{x},\mathbf{y})=-\sum_{m=1}^{M} x_m\log(y_m)+(1-x_m)\log(1-y_m) \tag{6}$$

が用いられることも多く，様々な関数が状況に応じて使い分けられる．

エンコーダ関数とデコーダ関数にニューラルネットワークを用いた場合，最適化は一般的なニューラルネットワークと同様に確率的勾配降下法によって行うことができる．オートエンコーダによる特徴抽出は，最適化によって得られたエンコーダ関数によって，データを特徴量に変換することで行われる．

2.2 オートエンコーダとモード分解

このような特徴抽出が一般的な枠組みであることを示すため，非定常流れの解析ツールである固有直交分解 (POD) がオートエンコーダとみなせることを示す．POD は少数の POD モードと呼ばれる場の線形和によって，オリジナルの場を近似する方法である[8),9)]．非定常流れの解析における強力なツールとして認識され始めた[10)]．

POD は，平均場を表すモードを φ_0，時間変動分を表すモードを $\varphi_k\,(k=1,\dots K)$ として，

$$\mathbf{x}_t\simeq\varphi_0+\sum_{k=1}^{K}\varphi_k\varphi_k^{\mathrm{T}}(\mathbf{x}_t-\varphi_0) \tag{7}$$

によって，オリジナルの場 $\mathbf{x}$ を近似する．この近似の誤差を式(5)の 2 乗誤差で測り，その総和

$$\sum_{t=1}^{N}\left|\mathbf{x}_t-\left(\varphi_0+\sum_{k=1}^{K}\varphi_k\varphi_k^{\mathrm{T}}(\mathbf{x}_t-\varphi_0)\right)\right|^2 \tag{8}$$

を最小化する線形独立なモード φ_k を見つける方法である．非定常流れの分析において，流れの変動の性質を捉えるものとして，それぞれのモードを可視化して用いる．

この手法は，エンコーダ関数を

$$\mathbf{h}=\mathrm{Encoder}(\mathbf{x})=(\varphi_1^{\mathrm{T}}(\mathbf{x}-\varphi_0),\dots,\varphi_K^{\mathrm{T}}(\mathbf{x}-\varphi_0)) \tag{9}$$

デコーダ関数を

$$\mathbf{y}=\mathrm{Decoder}(\mathbf{h})=\varphi_0+\sum_{k=1}^{K}\varphi_k h_k \tag{10}$$

と見れば，φ_k をパラメータとしたオートエンコーダとみなせる．この時の特徴量 $\mathbf{h}$ は，各瞬間の変動分 $\mathbf{x}_t-\varphi_0$ を各モードに正射影したベクトルの大きさであると解釈できる．

実際 POD は機械学習で古くから用いられている主成分分析 (PCA) と同値であることが知られている．さらに，復元誤差について 2 乗誤差を用い，エンコーダ関数とデコーダ関数を線形関数にしたオートエンコーダは，PCA と同様の部分空間を学習することが示されている．

2.3 オートエンコーダの盛衰

前項で見たように，少ない情報量でデータを復元できることが良いという思想による特徴抽出は，特段新しいものではない．エンコーダ関数やデコーダ関数にニューラルネットワークを用いた構造に限っても，1990 年代には auto-association と呼ばれて盛んに研究がなされた[11)]．しかしながら，シンプルなモデルでは取得できる特徴が乏しい一方で，複雑なモデルでの学習が困難であったことから，研究は下火になっていった．

しかし，2006 年 Hinton らによって多層オートエンコーダの学習が成功し[6)]，状況は一変した．これは，現在の深層学習の元祖となるモデルとも言え，ニューラルネットワークの研究を加速させた．彼らは各層を貪欲法的に学習する，すなわちモデル全体を一度に学習する代わりに個別に 1 層ずつ学習することで複雑なモデルの学習を達成した．さらに，2012 年には，画像から猫や顔といった高度な認識を示すニューロンが存在することが発見され，複雑な特徴量を学習できていることが示された[12)]．多層オートエンコーダによる特徴量の抽出法の発見は，ニューラルネットワーク，ひいては人工知能の発達に大きく寄与した．

一方で，一般にニューラルネットワークの構造は複雑化，巨大化している．モデルが複雑すぎるため，内部的に何が行われているかは分からず，解釈が難しいという問題が生じている．特にオートエンコーダは，特徴量を得ることが目的であるため，結果として何が捉えられているのか理解することが困難であることは現状の大きな

課題であるといえる．

3．オートエンコーダの種類

本節では，オートエンコーダを構成する際に，その構造に導入される構造やトリックについて，いくつか紹介する．これらの構造は，オートエンコーダの損失関数を変更することで過学習を防ぎ，汎化性能を高める役割を担っている．前節ではオートエンコーダは特徴量の次元を削減すると述べたが，これらの構造を用いれば過完備な（次元の高い）特徴量も取得できる．実際にオートエンコーダを構成する場合には，目的に応じてこれらの構造を適宜選択して利用することが多い．

3.1 スパースオートエンコーダ

特徴量の値の中に0が多い表現は「スパース」であると言われる．スパースな表現を獲得することは，スパースモデリングと呼ばれ，解釈性が高まるなど多くの利点があることが知られている．スパースオートエンコーダ[13]は，このスパースな特徴量を獲得するための構造である．

スパースオートエンコーダでは，損失関数を

$$L=\sum_{t=1}^{N} E(\mathbf{x}_t,\mathbf{y}_t)+\lambda \mathrm{S}(\mathbf{h}_t) \tag{11}$$

のように変更する．ここで，特徴量の関数 $\mathrm{S}(\mathbf{h})$ はスパース正則化項と呼ばれ，特徴量がスパースでない分に加えられるペナルティと解釈できる．λ は正則化の強さを表すパラメータである．この関数には，目標とするスパース率 ρ と実際のスパース率 $\overline{h_t}$ のとの差：

$$\mathrm{S}(\mathbf{h}_t)=-\rho\log\overline{h_t}-(1-\rho)\log(1-\overline{h_t}) \tag{12}$$

が広く用いられている．

3.2 デノイジングオートエンコーダ

デノイジングオートエンコーダ[14]は，人工的に入力にノイズを乗せて，ノイズのないオリジナルデータを復元するように学習させる．これによって入力が少々変化したとしてもロバストな変換の学習がおこなわれると期待される．

オリジナルのデータ $\mathbf{x}_t$ に対して，ノイズを乗せる操作を用意する．この操作には，平均0，分散一定の正規分布に従う値を一様に加えるガウスノイズ，データの中からランダムに抽出された要素の値を最大値または最小値にするソルト＆ペッパーノイズ，ランダムに抽出されたデータ点の値を0にするマスクノイズなどが用いられる．この操作によってノイズを加えられたデータ $\widehat{\mathbf{x}_t}$ をエンコード，デコードして得られる復元結果 $\widehat{\mathbf{y}_t}$ とする．デノイジングオートエンコーダでは，ノイズ付きデータからの復元結果 $\widehat{\mathbf{y}_t}$ とオリジナルの $\mathbf{x}_t$ との誤差 $\mathrm{E}(\mathbf{x}_t,\widehat{\mathbf{y}_t})$ が最小になるように学習を行う．

3.3 コントラクティブオートエンコーダ

入力の変化にロバストな変換を得たいというモチベーションから提案された別の構造として，コントラクティブオートエンコーダ[15]がある．この目的のためには，エンコーダ自体が急激に変動をおこさない関数であるべきである．そこでエンコーダ関数のヤコビアンの大きさに対してペナルティを与える．データ点の周りでのヤコビアンのフロベニウスノルムを加えたものを損失関数として用いる．

3.4 変分オートエンコーダ

オートエンコーダに確率モデルとしての構造を導入し，生成モデルとして扱うことができるようにしたものが変分オートエンコーダ（VAE）[16]である．特徴量を確率モデルにおける潜在変数とみて，ベイズ的アプローチによってモデルを構成する．特徴量もとい潜在変数を確率的にサンプルしてデコードすることで，今までに得たデータと類似のデータを生成することができる．類似の画像の生成などに応用されており，人工知能研究者の間での最近のホットトピックである．

VAEでは，エンコーダ関数が出力する特徴量の分布が，何らかの特定の確率分布に従うようにオートエンコーダを学習する．白色のガウス分布 $\mathcal{N}(0,\mathrm{I})$ に従わせる場合の損失関数は，

$$L=\sum_{t=1}^{N}\mathrm{E}(\mathbf{x}_t,\mathbf{y}_t)+\mathrm{KL}(\mathrm{P}(\mathbf{h}_t|\mathbf{x}_t)\|\mathcal{N}(0,\mathrm{I})) \tag{13}$$

のようになる．ここで，$\mathrm{KL}(\cdot\|\cdot)$ はKLダイバージェンスと呼ばれる確率分布間の違いを測る関数である．この損失関数は変分ベイズ法を用いて理論的に導出されるが，直観的にも特徴量が従ってほしい分布との乖離をペナルティとして加えていると解釈できる．

4．オートエンコーダを用いた分析例

著者らはオートエンコーダを用いて，非定常流れ場を分析する研究を行っている．2次元翼型を過ぎる流れの速度場データに対して，オートエンコーダを用いて時間変化の可視化を行った事例について述べる．

4.1 用いた流れのデータ

本稿で述べる分析に用いた流れのデータは，CFD計算による2次元翼型後方の非定常流れデータである．翼型にはNACA0012を選択し，Fig. 1に示すようなO型格子を用い，文献[17]の方法で計算を行った．レイノルズ数を $Re=1000\sim30000$ の間で7パターン，迎角を $\alpha=0^\circ\sim12^\circ$ の間で1°刻みの13パターン，合計91パターンの流れについて計算を行った．$t=0$ の静止状態から無次元時間 $t=5$ まで加速させ，その後は一定速度で動かし，$t=100$ まで計算を行った．それぞれの流れについて，Fig. 1に赤点で示した 40×40 の矩形格子状のサンプル点における速度を $\Delta t=0.025$ おきに計4000回記録し，データとして用いた．

本稿では，$Re=10000$，$\alpha=10^\circ$ の流れを対象に行った可視化の結果を示す．時刻AとBは，計算開始時点 $t=0$ の静止状態から徐々に翼を加速していく状態中の流れである．この流れでは，$t=5$ の加速終了後も初期渦は発達し続け，揚力は上昇する．時刻Cでは，サン

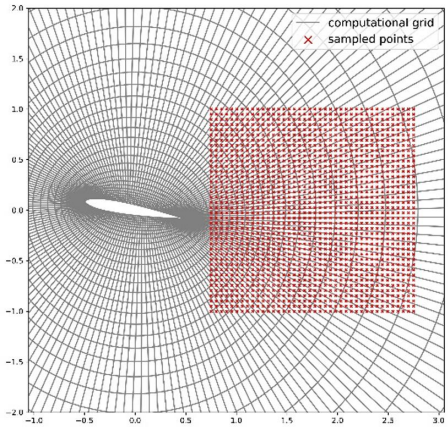

Fig. 1 Computational grid and the sampling points.

プル点の領域に初期渦が現れる．その後は周期的に渦が発生するようになり，時刻 D から F は，$t=75$ 近傍での周期構造の約 1 周期分について 3 つの時刻をとって表したものである．

4.2 オートエンコーダの構成

本研究では，エンコーダ関数とデコーダ関数の両者に畳み込みニューラルネットワークを用いたオートエンコーダを構成した．

入力と出力はサンプル点に相当する 40×40 点に対し，速度 (u,v) の 2 次元のデータが乗った，$40\times40\times2=3200$ 次元である．特徴量の次元は 200 次元とした．エンコーダ関数とデコーダ関数はそれぞれ畳み込みニューラルネットワークで構成し，活性化関数には ReLU 関数を用いた．

オートエンコーダの学習には 4.1 節で取得したすべてのデータ，すなわちすべてのレイノルズ数・迎角・時刻に対する速度場のデータを用いた．誤差関数 E として，L1 ノルムを用いた．

学習されたオートエンコーダを用いて，いくつかの流れ場に対して，特徴量を取得した．さらに，取得された特徴量をデコーダ関数に入れることで復元された流れ場を取得した．レイノルズ数 $Re=10000$，迎角 $\alpha=10°$ の流れにおける時刻 D と E の結果を Fig. 2 に示す．流れ場は速度ベクトルを矢印表示した．この図から，復元された流れ場と元の流れ場に大きな違いは見られない．実際，すべての流れ場に対する速度の平均誤差は，主流速の 0.96% であった．したがって，この特徴量は低次元量で流れ場を表現できていると考えられる．

4.3 特徴量を用いた時間変化の可視化

Fig. 2 に示した 200 次元の特徴量は，複数の流れ場に対して比較することは難しい．そこで，この特徴量をさらに主成分分析することで，2 次元の主成分スコアに変換し，これをプロットすることで可視化する．時間的に連続しているデータの特徴量を連続的にプロットすることで，点の移動軌跡としてデータの時間変化を表すことができる[18]．表示画面上の点が，ある時刻の瞬間での流れの空間構造を代表することになる．

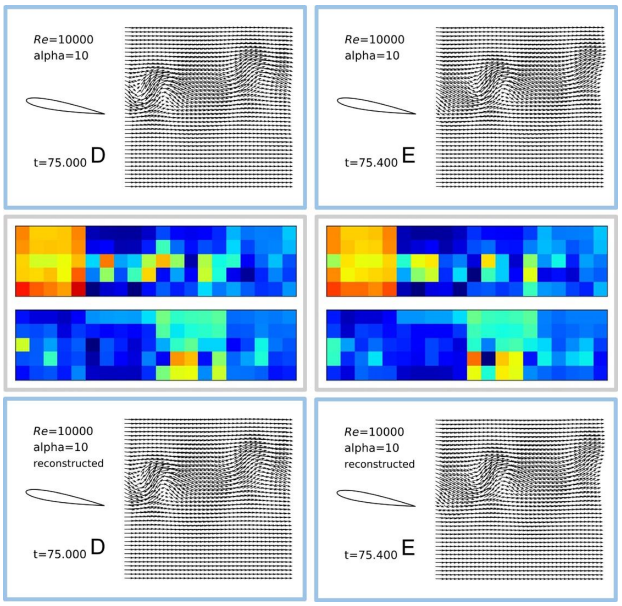

Fig. 2 Example of the result of the auto-encoder. Original flow fields (top), features obtained (middle), and reconstructed flow fields (bottom).

結果を Fig. 3 に示す．横軸が PCA における第 1 主成分，縦軸が第 2 主成分を表す．図中の各点がある時刻での流れ場の様子を表している．また，色によって時刻を表し，青から赤に向かって時間が進行する．図の右側には，各時刻 A から F における流れ場を表示した．さらに，Fig. 4 に翼の揚力係数の時系列グラフを示す．

提案手法による，低次元表現を用いた可視化の特徴について述べる．Fig. 3 右に示す，従来の速度ベクトルの可視化によって瞬間の流れの様子は読み取れる．しかしながら，このような瞬間場を取り出しただけでは流れの時間変化を捉えることは難しい．例えば，周期性の存在の有無を確認することは難しい．このため，Fig. 4 に示す揚力係数の時系列グラフを参照しながら，流れの時間変化を推測するという方法がとられることが多い．しかし，そのような方法は，空間構造の時間変化を直接的に示すものではない．一方，Fig. 3 左の提案手法による低次元特徴量の可視化表現では，流れ場の時空間構造が周期的な構造に漸近しながら完全に周期的な構造に移行することを示すことができる．

5. おわりに

本稿では，人工知能分野と可視化分野における低次元特徴量の必要性という共通点を指摘した上で，オートエンコーダによる特徴抽出法について概説した．さらに，可視化における応用例として，オートエンコーダを非定常流れの時間変化の可視化に用いた研究について述べた．今後も両分野において，新しいオートエンコーダの構造利用法についても発展していくことが期待される．

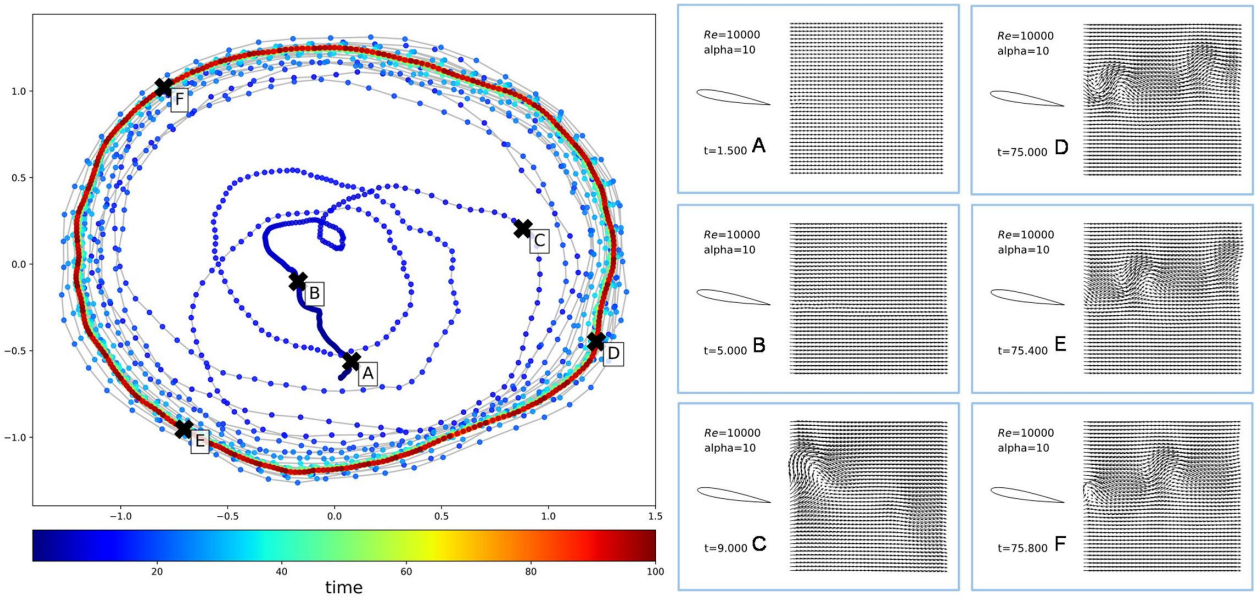

Fig. 3 Visualization of spatio-temporal structure of flow field through auto-encoder.

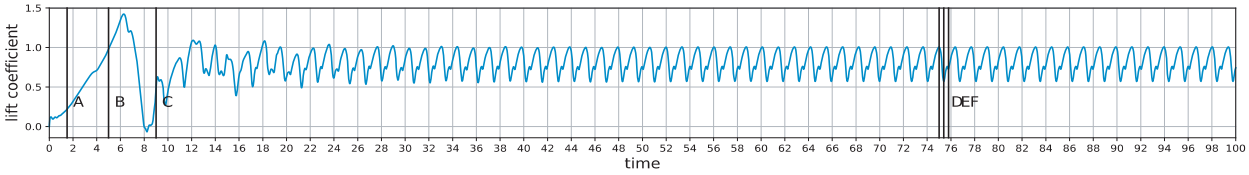

Fig. 4 Time series graph of the lift coefficient.

参 考 文 献

1）白山晋：知的可視化，丸善（2006）.
2）高間康史：情報可視化，森北出版（2017）.
3）坂野鋭，山田敬嗣：怪奇!! 次元の呪い -識別問題，パターン認識，データマイニングの初心者のために-（前編），情報処理，Vol.43, No.5（2002）, pp.562-567.
4）坂野鋭，山田敬嗣：怪奇!! 次元の呪い -識別問題，パターン認識，データマイニングの初心者のために-（後編），情報処理，Vol.43, No.6（2002）, pp.658-663.
5）山下隆義：統計的学習手法を用いた物体認識における特徴量の進化，画像ラボ，Vol.20, No.1（2009）, pp.53-58.
6）Hinton, G. E., Salakhutdinov, R. R.: Reducing the dimensionality of data with neural networks, Science, Vol.313 No.5786（2006）, pp.504-507.
7）Bengio, Y. et al.: Representation learning: A review and new perspectives, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.35, No.8（2013）, pp.1798-1828.
8）平邦彦：固有直交分解による流体解析：1．基礎，ながれ，Vol.30, No.2（2011）, pp.115-123.
9）平邦彦：固有直交分解による流体解析：2．応用，ながれ，Vol.30, No.3（2011）, pp.263-271.
10）Taira, K. et al.: Modal analysis of fluid flows: An overview, AIAA J., Vol.55, No.12（2017）, pp.4013-4041.
11）Kramer, M. A.: Nonlinear principal component analysis using autoassociative neural networks, AIChE J., Vol.37, No.2（1991）, pp.233-243.
12）Le, Q. V.: Building high-level features using large scale unsupervised learning, In Proc. of IEEE Int. Conf. Acoustics, Speech and Signal Processing（ICASSP）,（2013）, pp.8595-8598.
13）Ranzato, M. et al.: Sparse feature learning for deep belief networks, Advances in Neural Information Processing Systems（NIPS）, Vol.20（2008）, pp.1185-1192.
14）Vincent, P. et al.: Extracting and composing robust features with denoising autoencoders, In Proc. of the 25th Int. Conf. Machine Learning（ICML）,（2008）, pp.1096-1103.
15）Rifai, S., et al.: Contractive auto-encoders: Explicit invariance during feature extraction. In Proc. of the 28th Int. Conf. Machine Learning（ICML）,（2011）, pp.833-840.
16）Kingma, D. P., Welling, M.: Auto-encoding variational Bayes, In Proc. of Int. Conf. Learning Representations,（2013）, pp.8595-8598.
17）Shirayama, S.: Flow past a sphere: topological transitions of the vorticity fields, AIAA J., Vol.30, No.2（1992）, pp.349-358.
18）Van den Elzen, S., et al.: Reducing snapshots to points: A visual analytics approach to dynamic network exploration, IEEE Trans. Visualization and Computer Graphics, Vol.22, No.1（2016）, pp.1-10.