

時系列パターン監視による自動車応答特性の変化検知

Detection of Characteristics Changes of Automobiles with Time-Series Patterns

武石直也 *1
Naoya Takeishi

矢入健久 *2
Takehisa Yairi

佐藤守一 *3
Shuichi Sato

田川貴章 *3
Takaaki Tagawa

*1 東京大学大学院 工学系研究科 航空宇宙工学専攻
Department of Aeronautics and Astronautics, The University of Tokyo

*2 東京大学 先端科学技術研究センター
RCAST, The University of Tokyo

*3 株式会社豊田中央研究所
TOYOTA CENTRAL R&D Labs., Inc.

In this study, we discuss detection of automobile characteristics changes by monitoring the time-series patterns of sensor data. We propose a method based on sparse coding, dictionary learning and latent semantics analysis (LSA) to capture the patterns of multivariate time-series and their co-occurrences. The proposed method has been applied to the vehicle sensor data, and succeeded in detecting changes with better performance than traditional PCA-based approach that ignores temporal correlation.

1. 背景

動的システムである自動車の特性は、走行モードや走行環境の違いによって様々に変化する。このような特性の変化は、加速感の変化などといった形で運転手が感じることもあるが、センサデータから検出することができれば客観的な指標として有用である。しかし、データが次々と蓄積されていく中で、将来的に専門家によるきめ細かいチェックの負担増が懸念される。そこで、自動的にセンサデータをチェックし、特性の変化の有無を判定するようなシステムが望ましい。センサデータからの変化検出の一つとして、専門家があらかじめ基準状態のモデルを構築し、それにデータを当てはめて変化を検出する方法が考えられる。ところが、センサデータの種類や背後にあるメカニズムは車種や型式によって大きく異なるため、ある車種のモデルを他の車種のデータに適用するのは適切ではない。

そこで本研究では、基準状態として定義する「参照データ」とそれ以外の「テストデータ」とを比較し、テストデータ中に表れる自動車の特性変化を検知することを目標とする。このような枠組みでは、専門家による手動のモデル構築が不要であり、低コストな変化検出システムの構築が可能である。ここで問題となるのが、自動車センサデータの性質である。自動車センサデータは、車速、舵角、スロットル開度、空燃比、ギア位置などの多様な変数からなる異種混合データであり、いくつかの変数は強い相関を有する。また、走行環境の変化や運転手の操作によって無限ともいえるモードを呈し、極めて非定常である。このような多変数・異種混合・強相関・非定常の時系列データの解析は困難な問題であり、決定的といえる解析手法は存在しない。

時系列データからの変化検知や異常検知では、時間方向の相関を無視して各データ点を i.i.d. として扱う手法の有効性が複数示されている。例えば、人工衛星のテレメトリデータからの異常検知に対しては、確率的主成分分析 (PPCA) の混合モデルである MPPCA [Tipping 99] の有効性が確認されている [葱 09]。しかし、例えば加速感の変化を捉えるためには、各データ点を i.i.d. で扱うのではなく、時間方向の相関を保存した形でデータを扱うべきである。時間方向の相関を捉える方法

のひとつとしては、自己回帰モデルや状態空間モデルなどによる生成モデルの学習が挙げられる。異常検知への適用としては、自己回帰モデルとカーネルマシンを用いた適応的なリミットチェックがある [Fujimaki 05]。しかし、非定常性の強いデータをひとつまたは数個の生成モデルによって表すことは困難である。

本研究の方向性は、時系列データの「パターン」を監視することで、時系列データに表れる自動車の特性変化を検知するというものである。すなわち、参照データ中には見られないパターンがテストデータ中に見つければ、それは何らかの変化を表していると考えられる。パターンを監視する方法では、データの時系列性をよく捉えることができる。また、過度に複雑になり得る生成モデルを考えないので、よりコンパクトな学習方法・表現を用いることができる。以下では、時系列パターンからの変化検知手法に関して、第一段階として単変数時系列の監視を、第二段階として多変数時系列の監視を検討する。

2. 単変数時系列パターンの変化検知

時系列パターンの変化検知の第一段階として、変数系列ごとのパターンをそれぞれ監視する問題を考える。ここでは、異なる系列間 (例えば、車速とアクセル開度) の関係は考えず、それぞれの単変数時系列データを参照データと比較する。そのため、複数系列の関係性が変化しているような状態を検出することはできない。単変数時系列パターンの変化検知の方法として、最近傍法、特異スペクトル分解、及び疎表現と辞書学習を検討する。

なお、本研究全体を通して、滑走窓によって元の時系列データを部分時系列の集合に変換して扱う。すなわち、ある変数系列の時系列データ (長さ N) を $T = t_1, \dots, t_m, \dots, t_N$ としたとき、 $t = m$ ($m = 1, \dots, M$) から始まる部分時系列として $s_m = [t_m, \dots, t_{m+w-1}]^T \in \mathcal{R}^w$ と取る。ただし、 w は滑走窓サイズで、 $M = N - w + 1$ である。このような部分時系列をまとめて、ハンケル行列 $X = [s_1 \cdots s_M]^T \in \mathcal{R}^{M \times w}$ として表す。ただし、この X は一つの変数系列の時系列データに対応するものであり、ほんらいの複数系列のデータではないことに注意されたい。また、以下の表記において、参照データを添字 r で表し、テスト (クエリ) データを添字 q で表す。

2.1 最近傍法

単純な方法として、テストデータ中の部分時系列 s_m^q に関して、参照データ中の最近傍部分時系列 $s_{m'}^r$ を見つけることによってパターンの変化を検知する。すなわち、最近傍の部分時系列との類似度が低い部分は、参照データからの変化が大きいとみなす。ここでは、変化度スコア $C_m(m = 1, \dots, M^q)$ として、(1) 式のように最近傍とのユークリッド距離をとる。最近傍を探索する方法では、テストデータ長の 2 乗に比例して類似度計算の回数が増加し、計算コストは大きい。

$$\begin{aligned} C_m &= \|s_m^q - s_{m'}^r\|_2 \\ &= \sqrt{(t_m^q - t_{m'}^r)^2 + \dots + (t_{m+w-1}^q - t_{m'+w-1}^r)^2} \quad (1) \end{aligned}$$

2.2 特異スペクトル分解

特異スペクトル分解 (SSA) は信号処理の分野で用いられてきた手法で、Moskvina と Zhigljavskaya によって変化点検知への応用が提案されている [Moskvina 03]。訓練時には、参照データのハンケル行列 X^r の分散共分散行列に対して固有値分解を行う。固有値を λ_i 、固有ベクトルを $V = [\dots v_i \dots] \in \mathcal{R}^{w \times w}$ ($i = 1, \dots, w$) とし、 λ_i のうち大きいものから k 個を選び、それに対応する固有ベクトルを $V_k = [v_1 \dots v_k] \in \mathcal{R}^{w \times k}$ とする。

一方テスト時には、 V_k を用いてテストデータの再構成誤差を求める。テストデータ X^q の低次元への埋め込みは $X_{emb}^q = X^q V_k \in \mathcal{R}^{M^q \times k}$ となり、元の空間への再構成は $\hat{X}^q = X_{emb}^q V_k^T$ で与えられる。さらに、変化度スコア $C_m(m = 1, \dots, M^q)$ として、再構成誤差 $\tilde{X}^q = \hat{X}^q - X^q$ の全列に関する合計をとる。

$$C_m = \sqrt{\sum_{i=1}^w (\tilde{X}^q(m, i))^2} \quad (2)$$

2.3 疎表現と辞書学習

時系列パターンをとらえる方法として、疎表現 (sparse coding) と辞書学習 (dictionary learning) を用いる。この枠組みでは、データ X を表現するために、ほとんどの要素がゼロである疎表現ベクトル $\alpha \in \mathcal{R}^{M \times K}$ と、 K 個の基底パターンによって構成される辞書行列 $D \in \mathcal{R}^{K \times w}$ を学習する。 α と D は、(3) 式の最小化によって学習される。

$$\underset{\alpha, D}{\text{minimize}} \quad \|X - \alpha D\|_2 + \lambda \|\alpha\|_p \quad (p \in \{0, 1\}) \quad (3)$$

(3) 式の最適化は、 α に関する最適化と D に関する最適化を交互に反復して行う。本研究では、 α の最適化として、LARS アルゴリズム [Efron 04] を用いて L1 正則化 ($p = 1$) 最小化問題を解く。また、 D の最適化として、Mairal らの提案したアルゴリズム [Mairal 10] を用いる。

訓練時には、(3) 式の最適化を α と D 両方に関して行い、参照データの辞書 D^r を得る。テスト時には、辞書を D^r に固定して、 α に関してのみ最適化を行う。テストデータ中に参照データにはないパターンが含まれていた場合、その部分は参照データの辞書 D^r ではよく表現できないと考えられる。そこで α^q と D^r によるテストデータの再構成 $\hat{X}^q = \alpha^q D^r$ を用いて、再構成誤差 $\tilde{X}^q = \hat{X}^q - X^q$ を計算し、(2) 式と同様に、全列に関する合計を変化度スコア $C_m(m = 1, \dots, M^q)$ とする。

3. 多変数時系列パターンの変化検知

前節の手法では、多変数時系列データのそれぞれの変数を個別に監視しているため、変数間の関係性の変化をとらえることはできない。そこで、時系列パターン変化検知の第二段階として、複数変数の関係性を監視する問題を考える。ここでは、2.3 節で求めたデータの疎表現 α に対して、潜在意味解析によってパターンの共起関係を監視する。

3.1 疎表現による特徴抽出

2.3 節の手法において求めた α の非ゼロ要素は、各基底パターンの寄与度合いを表しており、元の時系列データを基底パターンの組み合わせで表した特徴量といえる。そこで、変数系列間との関係性を監視するために、 L 個の複数変数系列の疎表現 α_m^i ($i = 1, \dots, L$) を結合した大きな疎ベクトル

$$A_m = [\alpha_m^1 \dots \alpha_m^L]^T \in \mathcal{R}^{KL} \quad (m = 1, \dots, M)$$

を全時刻に関して作る。この大きな疎ベクトル A_m は、時刻 m の全変数にわたるパターンをひとつのベクトルでとらえた特徴量であり、当然ながら疎であって各非ゼロ要素はそれぞれの変数における基底パターンの寄与度合いを表している。

3.2 潜在意味解析

前節のような大きな疎ベクトル A_m を全時刻 $m = 1, \dots, M$ に関して並べた行列 $A = [A_1 \dots A_M]^T \in \mathcal{R}^{M \times KL}$ を考える。この行列 A は、各行が文書、各列が単語に対応する文書単語行列と類似している。すなわち、各時刻の全変数に関する部分時系列が文書にあたり、その部分時系列をなす基底パターンが単語にあたるという考え方である。

このような多変数特徴行列 (文書単語行列) A に対し、潜在意味解析 (LSA) を適用することで変数間の共起関係ををとらえて変化検知を行う。潜在意味解析では、文書単語行列の特異値分解 $A = U \Sigma V^T$ を行う。ここで、 U の各行は概念空間における文書の特徴ベクトルを、 V の各行は概念空間における単語の特徴ベクトルを表している。更に、上位 d 個の特異値とそれに対応する特異ベクトルのみを用いて $A_d = U_k \Sigma_k V_k^T$ のように低ランク近似を行うことで、本質的な意味だけを保存することができる。

訓練時には、参照データの特徴行列 A^r に対して特異値分解を行い、 U_k と Σ_k を保存する。テスト時には、テストデータの特徴行列の概念空間への変換 $A_{emb}^q = A^q V_k \Sigma_k^{-1} \in \mathcal{R}^{M^q \times k}$ を経て、再構成 $\hat{A}^q = A_{emb}^q \Sigma_k V_k^T$ を行い、再構成誤差 $\tilde{A}^q = \hat{A}^q - A^q$ を計算する。そして、系列 l に関する変化度スコア $C_m^l(m = 1, \dots, M^q, l = 1, \dots, L)$ として、

$$C_m^l = \sqrt{\sum_{i=(l-1)K+1}^{lK} (\tilde{A}^q(m, i))^2} \quad (4)$$

をとる。

4. 変化検知実験

4.1 実験の設定

単変数時系列パターン及び多変数時系列パターンの変化検知を実際の自動車センサデータに適用し、その有効性を確かめた。使用したデータは、自動車のエンジンコントロールユニット (ECU) で取得している 43 変数 (連続値 35 変数、離散値 8 変数) のセンサデータであり、車速、アクセル開度、エンジン回転数、前後左右の G、舵角、回生トルクなどの値からなる。本実験では、次のような走行状況で取得したデータを用いる。

データ 1: 基準 基準となるデータ。直線の道路を特別な状態なく走行。

データ 2: モード変化 運転モードをデータ 1 とは変更して走行。空燃比等の制御が変化しており、より加速しにくい状況になっている。

データ 3: 断続 N 直線の道路を、断続的にシフトをニュートラルに変更しながら走行。基準状態では減速時に回生トルクが発生するが、シフトがニュートラルに入っていると回生トルクが発生しないという変化がある。

データ 1 を基準データ、データ 2 及びデータ 3 をテストデータとして変化検知手法を適用し、期待される変化が検出できるか確かめる。データの特性から期待される実験結果は、データ 2 では加速時に空燃比に関係する変数の変化度スコアが高くなること、データ 3 では減速時に回生トルクに関係する変数の変化度スコアが高くなることである。

なお、滑走窓のサイズは $w = 8$ とした。また、辞書学習においては単語数 $K = 24$ とし、各 α の最大非ゼロ要素数は 4 個とした。特異スペクトル分解、及び潜在意味解析での次元削減では、固有値の累積寄与率が 90% を超える最小の値に k を設定した。

また、参考として、時系列性を考えない次元削減による変化検知についても実験した。時系列性を考えない次元削減では、43 変数の元データを主成分分析 (PCA) によって固有値の累積寄与率が 90% を超える最小の次元数に削減した後、系列ごとの再構成誤差を変化度スコアとした。

4.2 実験結果

2. 節及び 3. 節で述べた手法を適用した結果を以下に示す。なお、各グラフの (a)1 行目は走行状態把握のために表示している元データ (車速、エンジン回転数、空燃比、回生トルク) である。(b)2 行目は時系列性を考えない次元削減 (主成分分析) による変化度スコアで、(c)3 行目以降は 2. 節及び 3. 節で述べた手法による変化度スコアである。

4.2.1 データ 2: モード変化の実験結果

データ 2: モード変化に対する変化検知実験結果として、空燃比に関する変化度スコアを図 1 に示す。この例では、時系列性を考える全ての手法において加速時に変化度スコアが高くなる傾向があり、期待される結果と一致している。しかし、時系列性を考えない主成分分析では加速時以外に変化度スコアが高く出ており、期待される結果とは一致しない。

変化度スコアのしきい値を変えたときの ROC 曲線は図 2 のようになる。図 2 では、特異スペクトル分解の性能がわずかに優れている。また、潜在意味解析によって変数間の関係を考えたときとそうでないときの差は明確ではない。これは、モード変化による空燃比制御の変化が、他の変数との関係性というより空燃比のパターンそのものの変化に表れているためと考えられる。なお、ROC 曲線の描画時には、変化度スコアがしきい値以上であるときに車両が加速中である場合を正解 (True Positive) とした。

4.2.2 データ 3: 断続 N の実験結果

データ 3: 断続 N に対する変化検知実験結果として、回生トルクに関する変化度スコアを図 3 に示す。この例では、最近傍法、特異スペクトル分解、及び疎表現/辞書学習によって単変数ごとのパターンの変化を見る方法では変化度スコアが常にゼロである。しかし、時系列性を考えない主成分分析 (図 3(b)) や疎表現/辞書学習及び潜在意味解析 (図 3(f)) ではシ

フトがニュートラルのときに変化度スコアが高くなっており、期待される結果と一致している。

変化度スコアのしきい値を変えたときの ROC 曲線は図 4 のようになる。図 4 から、時系列性を考えない主成分分析よりも、疎表現/辞書学習+潜在意味解析により時系列パターンの共起関係を学習する手法の性能が良いことがわかる。なお、ROC 曲線の描画時には、変化度スコアがしきい値以上であるときにシフトがニュートラル、アクセル開度がゼロ、かつ減速中である場合を正解 (True Positive) とした。

5. 結論

本稿では、自動車センサデータのパターンを監視することで自動車特性の変化を検知する手法について検討した。単変数時系列の変化検知手法として、最近傍法、特異スペクトル分解、疎表現/辞書学習を適用した。また、多変数時系列の変化検知手法として、疎表現/辞書学習によって抽出したパターンの特徴量による行列に対して潜在意味解析を適用し、変数間関係の変化を検出する手法を提案し実験した。

実験の結果として、単変数のみでも変化が見られる事例では、時間方向の相関をとらえた方法によって期待される変化が検出されることを示した。また、複数変数間の関係性が変化している事例では、変数間のパターンの共起関係を学習することによって期待される変化がよく検出されることを示した。

本稿で提案した疎表現/辞書学習による特徴抽出と潜在意味解析による共起関係の学習は、きわめて単純な発想に基づくものであって、今後十分な検討と改良が必要である。例えば、疎表現と辞書を最適化する手法、各種パラメータの設定、または潜在意味解析以外の共起関係学習手法等については、今後の重要な課題である。

参考文献

- [Efron 04] Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R.: Least Angle Regression, *Annals of statistics*, Vol. 32, No. 2, pp. 407–499 (2004)
- [Fujimaki 05] Fujimaki, R., Yairi, T., and Machida, K.: Adaptive Limit-checking for Spacecraft using Relevance Vector Autoregressive Model, in *International Symposium on Artificial Intelligence, Robotics and Automation in Space* (2005)
- [Mairal 10] Mairal, J., Bach, F., Ponce, J., and Sapiro, G.: Online Learning for Matrix Factorization and Sparse Coding, *Journal of Machine Learning Research*, Vol. 11, pp. 19–60 (2010)
- [Moskvina 03] Moskvina, V. and Zhigljavskaya, A.: An Algorithm Based on Singular Spectrum Analysis for Change-Point Detection, *Communications in Statistics - Simulation and Computation*, Vol. 32, No. 2, pp. 319–352 (2003)
- [Tipping 99] Tipping, M. E. and Bishop, C. M.: Mixtures of Probabilistic Principal Component Analysis, *Neural Computation*, Vol. 11, No. 2, pp. 443–482 (1999)
- [葱 09] 葱 遞, 蛄蜈 蛛葱, 豊蟻毯蛄我, 逕逕 蜥蛄寇 壽蜈 炎 貂帙 蜀肴区 隱蟾繡堤烟緜 逕蝗讀勳衍香区輔 豈碑, 隨 23 蝗模蟾蛄間蟄莨壺 蝗蜈莨 (2009)

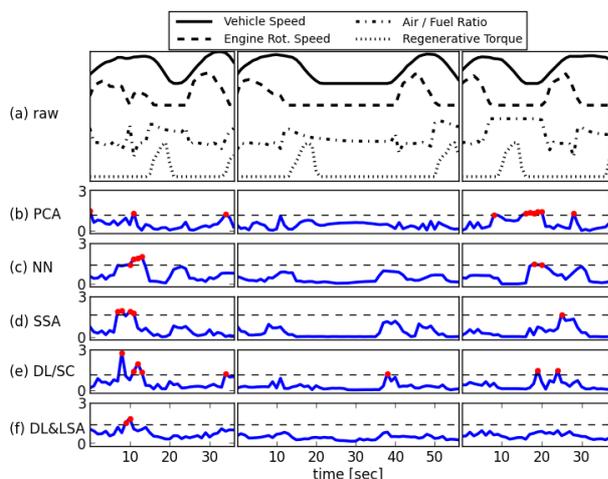


図 1: データ 2: モード変化における空燃比の変化度スコア。2 行目 (b): 時系列性を考えない PCA、3 行目 (c): 最近傍法、4 行目 (d): 特異スペクトル分解、5 行目 (e): 疎表現と辞書学習、6 行目 (f): 辞書学習+潜在意味解析。1 行目 (a) は元データである。赤マーカ部分は変化度がしきい値以上の部分。

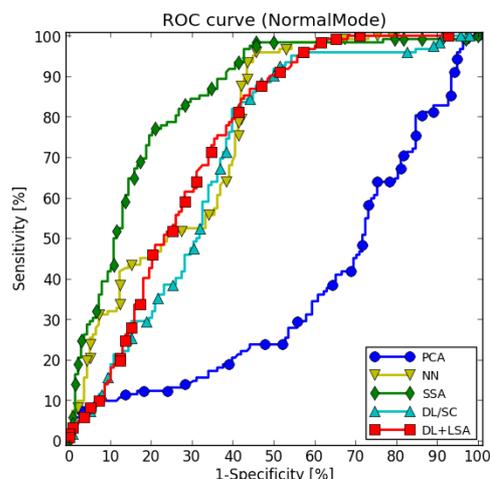


図 2: データ 2: モード変化における空燃比の変化検知についての ROC 曲線。変化度スコアがしきい値以上であるときに車両が加速中である場合を正解 (True Positive) とした。

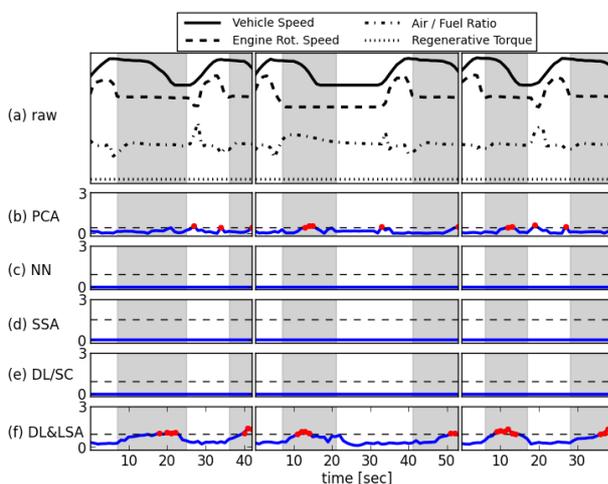


図 3: データ 3: 断続 N における回生トルクの変化度スコア。2 行目 (b): 時系列性を考えない PCA、3 行目 (c): 最近傍法、4 行目 (d): 特異スペクトル分解、5 行目 (e): 疎表現と辞書学習、6 行目 (f): 辞書学習+潜在意味解析。1 行目 (a) は元データである。赤マーカ部分は変化度がしきい値以上の部分。網かけ部分はシフトがニュートラルの期間。

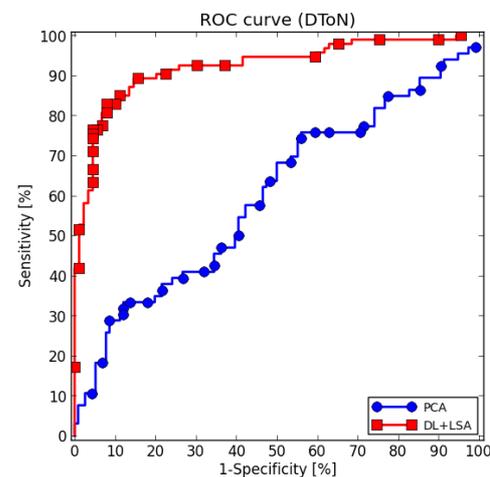


図 4: データ 3: 断続 N における回生トルクの変化検知についての ROC 曲線。変化度スコアがしきい値以上であるときにシフトがニュートラル、アクセル開度がゼロ、かつ減速中である場合を正解 (True Positive) とした。