

# 自動車用自動変速機の変速制御時に現れる時系列データの分類

Classification for time-sequence data appeared in shift control of automobile automatic transmission

森川祐介 \*<sup>1</sup>  
Yusuke Morikawa

石原靖弘 \*<sup>1</sup>  
Yasuhiro Ishihara

井手貴範 \*<sup>1</sup>  
Takanori Ide

森山英二 \*<sup>1</sup>  
Eiji Moriyama

秋田拓 \*<sup>1</sup>  
Taku Akita

田部井靖生 \*<sup>2</sup>  
Yasuo Tabei

宇津呂武仁 \*<sup>3</sup>  
Takehito Utsuro

中川裕志 \*<sup>2</sup>  
Hiroshi Nakagawa

\*<sup>1</sup>アイシン・エイ・ダブリュ(株)  
Aisin AW Co., Ltd.

\*<sup>2</sup>理化学研究所 革新知能統合研究センター  
RIKEN, Center for Advanced Intelligence Project

\*<sup>3</sup>筑波大学システム情報系  
Fclty. Eng. Inf. & Sys, Univ. of Tsukuba

The automatic transmission is a transmission that automatically changes gear according to an optimal gear step decided based on driver's accelerator/brake operation, vehicle speed, etc. Since the quality of a changing gear of automatic transmission, we call shift feeling, has a great influence on riding comfort of the automobile, the parameters of control program are adjusted to improve the shift feeling at the development site. We call this adjustment work calibration. In the process of calibration, there is a process of identifying occurring phenomena in the time-sequence data and selecting the cause and countermeasures. Because this task takes large man-hours, it is very efficient if this task becomes automatic. Therefore, at the first step, in order to automate identification of the occurring phenomena, we built classifier of the time-sequence data using SVM. The experimental result suggests that the proposed identification system is promising directions for actually used identification system.

## 1. はじめに

オートマチック車における自動変速機はユーザのアクセル操作やブレーキ操作、自動車の車速などの走行状態に応じて最適なギヤ段を決定し、自動的に変速を行う変速機である。ユーザの操作通りに変速するマニュアル変速機では変速時の空走感やショックは許容されるが、自動変速機はユーザの操作とは異なるタイミングで変速するために変速時のわずかな空走感やショックでも不快と感じる傾向がある。自動変速機の変速の質(変速フィーリングという)は自動車の乗り心地に大きな影響を与えており、開発現場では変速の質を高めるために自動変速の制御プログラムのパラメータ調整に大きな工数をかけている。技術者が判断した変速フィーリングが良くなるようにパラメータを調整することを適合と呼ぶ。

適合では以下の Step 1.~Step 4. を実施している。

Step 1. 自動変速機を自動車や評価機に搭載して動作させ、動作時の制御・物理データを時系列データとして取得する。

Step 2. 変速フィーリングに問題がある場合、時系列データを確認し、生じている問題の種類(=現象)を特定する。

Step 3. 特定した現象に対して想定される要因や対策の候補をいくつか抽出する。

Step 4. Step 3. の候補の中から、発生した現象の要因と対策を特定して制御プログラムのパラメータを調整する。

この一連の工程のうち Step 4. の要因と対策の特定には経験が必要であり、誤った要因と対策を選んでしまうと手戻りが発生する。そのため自動的に熟練者の判断に近い要因と対策を提示できれば、手戻りを未然に防止できる可能性が高くなり工数低減が見込める。また同時に、経験の少ない技術者が熟練技術

者の知見を得ることができるため、技術者の育成に繋がることも期待できる。

要因と対策の提示のためにはまず現象の特定が必要であるため、本研究では時系列データで発生した現象の特定を現象の分類問題として考え自動分類を行う。過去の研究 [1][2] では特徴量ベースで時系列データの分類を行っているが、本研究ではサンプリングした時系列データそのものを入力としてサポートベクターマシン (SVM) および K 近傍法 (KNN) を用いて分類する方法を検討した。

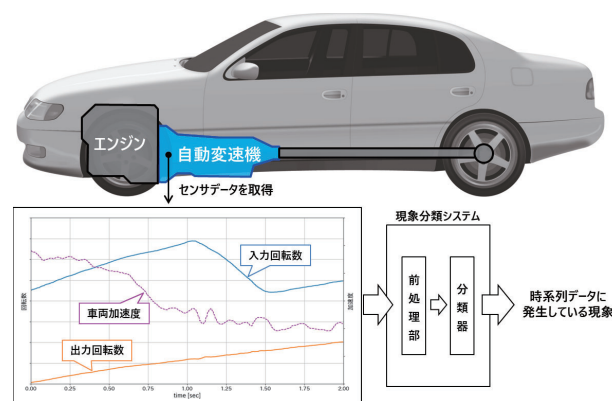


図 1: 自動変速機からの取得データと本研究のイメージ

## 2. 扱う時系列データ

適合時に取得した時系列データのうち、変速フィーリングに問題があるものを対象とする。時系列データに含まれる信号のうち回転数、加速度など 7 種類を用いた。時系列データは発

連絡先: 森川祐介, アイシン・エイ・ダブリュ(株), 〒444-1192  
愛知県安城市藤井町高根 10, i37710\_morikawa@aisin-  
aw.co.jp

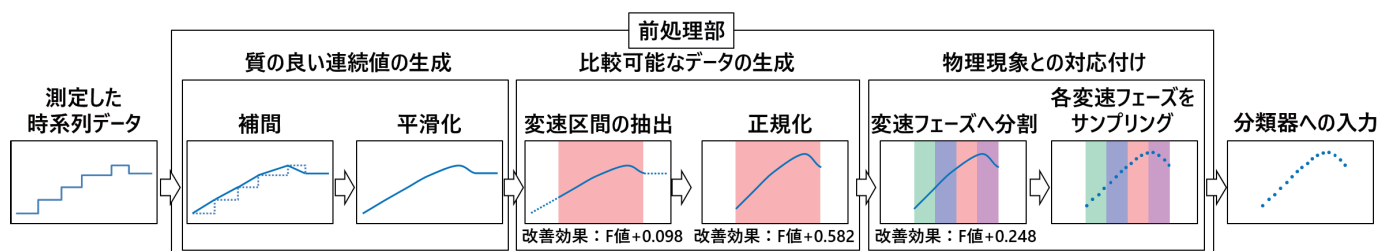


図 2: 前処理の流れ

生じた現象のラベルを持ち、1つのデータが複数のラベルを持つこともある。本来現象ラベルは複数の技術者により付けられているが、人によるばらつきを抑えるために1人の熟練技術者が特別に再度ラベル付けを行ったデータを用いた。

データベースに保存されている時系列データは長さ2秒、値の保存間隔は5ミリ秒、データ点数は400点である。検証のため、アクセルを踏み込みながら高速ギヤ段に変速するアップシフトのみを対象とする。時系列データは約300個あり、全24現象あるうちの数が多い順に7現象を対象とする。7現象に該当するデータ数は平均で44.6個、最大で86個、最小で16個である。

現象には、変速時にクラッチが勢いよく係合し、ショックが生じる現象などが含まれる。

### 3. 現象分類システムの構成

構築した現象分類システムの構成について説明する。現象分類システムは前処理部と分類器に分かれており、実データに適用するためには特に前処理部が重要となる。

#### 3.1 前処理部

図2に前処理の流れを示した。時系列データの7信号それぞれに対し前処理を行う。各前処理を順に説明する。

Step1. 時系列データの補間を行う。時系列データは5ミリ秒刻みで値が保存されているが、自動車の種類や開発フェーズ、取得環境により値の更新間隔が異なり、階段状になっていることがある。例えばあるデータでは5ミリ秒ごとに値が更新されるが、別のデータでは50ミリ秒ごとに値が更新される。そのため値が更新された点のみを用いて線形補間でデータを補間し、すべての時系列データを同じ時刻に比較出来るようにする。

Step2. 時系列データの平滑化を行う。目的はノイズ除去および線形補間で生じた不連続部の平滑化を行い、質の良い連続値を生成することである。移動平均期間は15期間とした。

Step3. 2秒間の時系列データから実際に変速している区間を抽出する。変速は0.5～1.5秒程度で終わり現象は変速中に発生するため、変速区間以外は現象の分類には必要ない情報であり、除外することで分類性能の向上が期待できる。

Step4. データのスケールを揃えるため、抽出した区間を正規化する。正規化手法には $z$ 正規化を用いた。時系列データ内の信号間で値のオーダーが数倍異なり、同じ信号でもデータ間で10倍程度値のオーダーが異なるため、正規化は必須である。

Step5. 抽出した区間を自動変速機の内部状態に応じた変速フェーズに分割する。変速フェーズは4つあり、順にクラッチをストロークさせるフェーズ、回転変化が起きるまでのフェーズ、回転が変化しているフェーズ、回転変化が終わり次のギヤ

段へ移行するフェーズである。現象は自動変速機内の物理現象が原因で発生するため変速フェーズと関わりが深く、特定の変速フェーズで特徴が出ることが多い。変速のたびにそれぞれの変速フェーズの長さが異なるため、変速フェーズに分割することで分類器の入力ベクトル内で現象の特徴が生じる位置を揃えることができる。実際、変速フェーズに分割したことで分類性能は大きく改善した。

Step6. 変速フェーズごとに時系列データをサンプリングし、時間長の異なるデータを同じ長さの入力ベクトルに変換する。サンプリングは、各変速フェーズを $N$ 個の小区間に分け、小区間内の最大値・最小値を抽出して行った。最大値・最小値を抽出する理由は、現象の判定に重要なピークやボトムを見落とさないようにするためである。今回は $N=10$ とした。

以上の前処理により、時系列データを長さ560の入力ベクトルに変換した。4.1節でそれぞれの前処理の効果を確認した。

#### 3.2 分類器

分類器にはSVMおよびKNNを用いた。SVMのカーネルは線形カーネルおよびRBFカーネルを検討した。分類は1対他分類で行ったため、データの正例・負例に不均衡が生じる。正例・負例それぞれでSVMの $C$ パラメータに各クラスのデータ数の逆数に比例する重みを掛け、不均衡データが学習に与える影響を抑制した。SVMの学習および評価は3-fold交差検証で行った。正則化パラメータ $C=10^{-3} \sim 10^3$ 、RBFカーネルパラメータ $\gamma=10^{-3} \sim 10^3$ を候補とし、交差検証の学習データを用い、2-fold交差検証でF値が最もよくなるパラメータを選択した。KNNは3-fold交差検証で評価し、近傍数 $k$ は $k=1, 3, 5, 7, 9$ を検討した。

### 4. 結果と考察

線形カーネルSVM (SVM(線形))、RBFカーネルSVM (SVM(RBF)) 及びKNNを用いて7現象の分類を実施した。

#### 4.1 前処理の効果

SVM(線形)、SVM(RBF)、KNNの中で最も良好な性能を示したSVM(RBF)において、各前処理のF値の改善効果を確認した(各分類器の性能については後述する)。システム上必要な処理を除いた、変速区間抽出、正規化、変速フェーズ分割を除外した場合を検討した。改善効果は、すべての前処理を実施したmacro-F値(0.711)と前処理を除外したmacro-F値の差として定義する。

変速区間抽出を外して2秒間の波形全体で正規化を行った場合、macro-F値は0.613となった。変速区間抽出の効果は+0.098である。現象は変速中に発生するため変速区間以外は現象とは関係のない無駄なデータであり、取り除くことで性能が向上したのだと考えられる。

正規化を外した場合、macro-F 値は 0.129 となった。正規化の効果は +0.582 である。現象ごとに交差検証の各 fold の結果を見ると、すべて正例または負例と判定されており分類器として機能していない状態であった。3.1 節で述べた通り信号間、データ間で値のオーダーが数桁異なるため、今回対象とする現象分類においては正規化は必須である。

変速フェーズ分割を外した場合、macro-F 値は 0.463 となった。変速フェーズ分割の効果は +0.248 である。前述した通り、現象は自動変速機内部の物理現象と密接な関係があるため、変速フェーズに分割することで入力ベクトル内で現象の特徴の位置が揃うようになる。そのため現象の分類に必要な入力が明確になり、性能向上に繋がったと考えられる。

## 4.2 KNN の結果

図 3 に  $k$  を変化させたときの F 値の推移を表示した。macro F 値は  $k=1$  がベスト性能、個別で見ても現象 3, 4 は  $k=3$  がベスト性能、その他現象は  $k=1$  がベスト性能となった。小さい  $k$  がベスト性能となる理由は、データが入り組んでおり境界面が引きづらいこと、正例の数が少ないことの 2 点だと考えられる。以下で詳細に確認する。

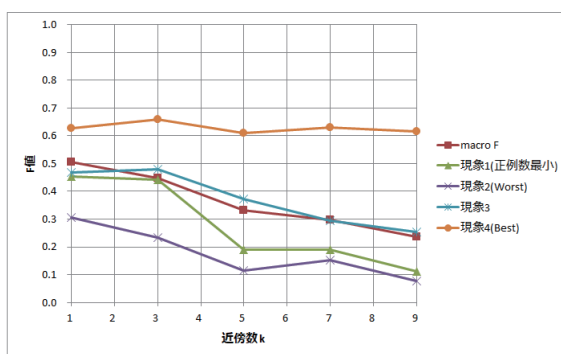


図 3:  $k$  を変更した際の KNN の分類性能

### 4.2.1 KNN ベスト性能

ベスト性能の現象 4 は正例数が 86 個と最も多く、また最後の変速フェーズでピークが立つため特徴が分かりやすい。さらに現象 4 と似た特徴を持つ現象が存在せず、後述する現象 2, 3 と比較して境界面が引きやすいと考えられる。しかし特徴は時系列データの一部にしか生じず入力ベクトルには現象の分類に不要な入力が含まれており、特徴空間上でデータは入り組んでいると考えられるため、大きな  $k$  では性能が出づらく  $k=3$  が最大となったと考えられる。

### 4.2.2 KNN ワorst性能

ワorst性能の現象 2 は正例数 39 個でデータ数が少ない。特徴は 2,3 番目の変速フェーズにまたがって発生し、後述する現象 3 と特徴が良く似ており技術者が見ても判別が難しい。特徴が良く似た現象 2, 3 でデータが入り組んでおり境界面が引きづらく、大きな  $k$  では性能が出ないのだと考えられる。

### 4.2.3 中間の性能

現象 3 は正例数 84 個で 7 現象中 2 番目に多いため、現象 2 より判定に使えるデータが多く性能も高くなっている。しかし  $k$  を大きくすると性能が落ちる傾向は変わらないことから、現象 2 との間でデータは入り組んでいると考えられる。現象 1 は正例数 16 個であるものの、似た現象が存在せず最初の変速フェーズにピークが出て特徴が分かりやすい。そのため  $k=1, 3$  ではデータ数のわりに分類性能が高くなっているが、

正例判定に学習データの半数が必要となる  $k=5$  以上では急激に性能が落ちている。

以上のことから、データが入り組んでおり境界面が引きづらいこと、正例の数が少ないことの 2 点により  $k=1, 3$  でベスト性能となったと考えられる。

## 4.3 KNN と SVM の比較

表 1 に SVM と KNN の評価結果を示した。KNN の  $k$  は現象ごとに最も F 値が良くなるものを選んだ。表に載せていない現象も含め、すべての現象で SVM の F 値が KNN の F 値を上回った。入力ベクトルには現象の分類に不要な入力も含まれているため、すべての入力を等価に扱う KNN では重みの調整で現象に関係する入力に注目できる SVM に性能が及ばないのだと考えられる。SVM(線形)の重みを確認すると、確かに人間が現象の判別を使う付近の入力は大きな絶対値の重みとなり、それ以外は 0 に近い値となっていることが多かった。

表 1: 分類結果

項目		SVM (線形)	SVM (RBF)	KNN
全体 (macro)	F-score	0.683	0.711	0.509
	Precision	0.588	0.661	0.608
	Recall	0.815	0.769	0.438
	AUC	0.869	0.876	-
現象 5	F-score	0.476	0.525	0.419
	Precision	0.374	0.439	0.493
	Recall	0.667	0.667	0.389
	AUC	0.787	0.778	-
現象 6	F-score	0.844	0.857	0.651
	Precision	0.781	0.915	0.761
	Recall	0.926	0.810	0.574
	AUC	0.977	0.942	-

## 4.4 SVM(線形) と SVM(RBF) の比較

すべての現象で SVM(RBF) の F 値が SVM(線形) の F 値を上回った。KNN で  $k=1, 3$  がベスト性能となったことをふまえると、特徴空間内ではデータ同士が複雑に入り組んだ構造になっており、非線形な境界面を引ける RBF カーネルが良好な分類性能を示したと考えられる。

## 4.5 SVM(RBF) の結果

### 4.5.1 正例・負例を判別できる現象

SVM(RBF) のベスト性能である現象 6 について、図 4 に一つの fold のテストデータの識別境界面からの距離の分布 (左) と ROC 曲線 (右) を示す。距離の分布から、正例 (上側) と負例 (下側) の分布がほぼ完全に分離しており正確に判別できていることがわかる。False Negative データは境界面に近い位置で間違えており、性質の良い境界面を生成できている。

現象 6 では変速の終盤に入力回転数の変動という特徴が現れ、他に似た現象が存在しない。同様に似た現象が存在しない現象 1, 4, 7 についても F 値が 0.8 前後、AUC 0.9 前後で精度よく判別できていることから、これらの現象では特徴空間内で現象ごとにデータが比較的まとまって分布していると考えられる。現象 6 の Recall が 0.810 に留まったのは、データ数が 26 個しかなく学習データでテストデータの事例をカバーしきれないことが原因だと考えられるため、データ数を増やすことで精度の向上が期待できる。



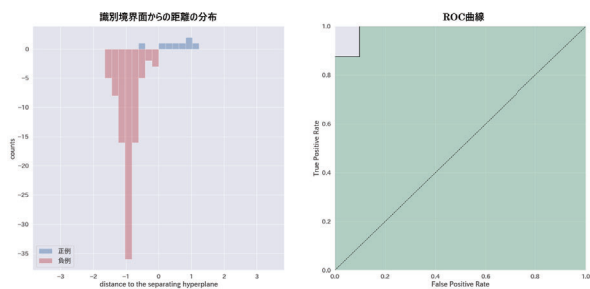


図 4: 現象 6 : 境界面からの距離の分布と ROC 曲線

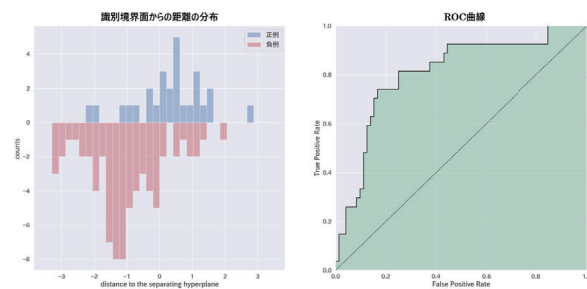


図 6: 現象 3 : 境界面からの距離の分布と ROC 曲線

#### 4.5.2 正例・負例の判別が難しい現象

ワースト性能の現象 5 についても同様に、図 5 に識別境界面からの距離の分布 (左) と ROC 曲線 (右) を示した。正例・負例の分布が相手の分布の深い位置まで食い込んでおり、SVM が分類する条件を学習できなかった現象が存在していることがわかる。False Positive データを確認すると、距離 1 付近のデータの約半数は現象 2 または現象 3 であった。また、距離 -3 付近の False Negative データでも、近い順に 10 個のデータを確認すると半数はやはり現象 2 または現象 3 であった。

現象 5 は現象 2, 3 と同じように 2, 3 番目の変速フェーズで特徴が生じる。特徴も現象 2, 3 と良く似ているため、今回分類に用いた入力だけでは現象を説明する情報が足りず、特徴空間内で重なって分布したのだと考えられる。

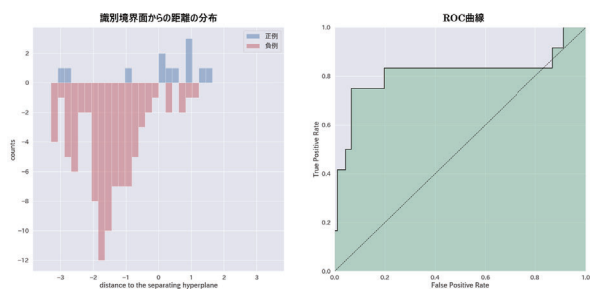


図 5: 現象 5 : 境界面からの距離の分布と ROC 曲線

図 6 に現象 3 の識別境界面からの距離の分布を示す。現象 5 同様、正例・負例の分布が重なっている。現象 3 の場合、False Positive となったデータの半数は現象 2 である。前に述べた通り現象 2, 3 は特徴が良く似ているため、現象を説明する情報が足りず特徴空間内でデータの分布が重なってしまったのだと考えられる。

現象 2 についても現象 3, 5 と同様の結果となった。良く似た現象を持つ現象 2, 3, 5 の分類性能を実用レベルまで上げるためには、特徴空間内で各現象の分布を切り分ける軸を与えられる新しい入力の検討が必要であると考えられる。

## 5. まとめと今後の課題

我々は、自動車用自動変速機の開発効率向上のため、変速フィーリングに問題がある時系列データに発生している現象の分類を行う手法を検討した。7 現象中 4 現象は現時点で実用レベルに近いところまで来ており、本手法で現象の分類は可能であるとする。しかし似た現象同士の分類精度は不十分である

ため、各現象を切り分けるための新しい入力の検討が必要である。

データ数と分類精度の関係について検討した結果を図 7 に示す。この図は 311 個の時系列データから無作為に  $N$  個を抽出して RBF カーネル SVM で学習・評価する試行をデータ数  $N$  ごとに 20 回行い、F 値の平均値を算出した結果である。データ数  $N$  は  $N = 63, 78, 105, 155, 206, 233, 248, 311$  とした。この図から、データ数の増加に伴い分類性能が向上する傾向が出ることがわかった。現象 1, 4, 6 については性能が飽和する傾向が出ているが、前述のとおりすでに十分な性能を得られているため問題はない。分類が難しい現象 2, 3, 5 はデータ数を増やすことで性能が向上しており、データ数を増やすことでも分類性能の改善が見込めることがわかったため、学習に利用できるデータを増やし分類器の改善を行っていく。

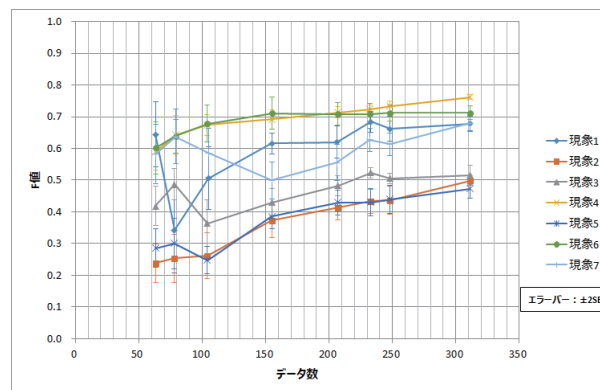


図 7: データ数と精度の関係

## 参考文献

- [1] B. D. Fulcher and N. S. Jones: Highly Comparative Feature-Based Time-Series Classification, IEEE Transactions on Knowledge and Data Engineering, Vol.26, No.12, 2014.
- [2] A. Kampouraki, G. Manis, and C. Nikou: Heartbeat Time Series Classification With Support Vector Machine, IEEE Transactions on Information Technology in Biomedicine, Vol.13, No.4, 2009.