

MDL 原理を用いた車両挙動データからの特徴抽出手法の提案

横山達也^{†1,a)} 白石陽^{†2,b)}

ドライバは道路上の交通状況によって快適な運転を妨げられることがある。例えば右折待ちをしている車列が走行経路上に存在する場合、その車列に接近する車両は非効率な減速や車線変更を強いられることがある。このような状況を回避するには、事前にドライバが走行予定経路上の交通状況（車群の挙動）を把握する必要があると考える。そこで本稿では、こうした交通状況を形成する車群の挙動を把握するための各車両の挙動推定に向けた特徴抽出手法を提案する。本研究では、車両挙動を、ドライバの運転行動の時系列から構成されるものとする。例えば右折の車両挙動は、車道の右側へ車を寄せ、減速し、ハンドルを右へ切るといった一連の運転行動の時系列の結果として成り立つと考えられる。運転行動の時系列性から車両挙動を推定することは、車群の挙動の把握に有用であるとする。本研究では、車載スマートフォンで収集したセンサデータに SAX(Symbolic Aggregate Approximation)を適用することで、時系列センサデータを文字列へ変換する。そして、自然言語処理技術である N-gram により、車両挙動を表す文字列から車両挙動の特徴を表す運転行動を部分文字列として抽出する。本稿では、MDL(Minimum Description Length)原理に基づいて、部分文字列の「時間長」と「出現頻度」に着目した特徴抽出手法の検討を行う。評価実験として、車載スマートフォンで収集した車両挙動に関するセンサデータに提案手法を適用することで、車両挙動の特徴を表す部分文字列の抽出を行った。その結果、「右折」、「右寄せ」、「減速」といった運転行動に該当する部分文字列が、特徴的な部分文字列として抽出することができた。

1. はじめに

近年、車が持つセンサ情報を、ネットワークを介して共有することにより、生活に役立つ情報を作り出すことを目的としたプローブ情報システムが実用化されている[1]。実用化されているサービスとして、ネットワークを自動車に接続することで利用できるテレマティクスサービスが挙げられ、T-Connect[2]やインターナビ[3]、NissanConnect[4]などが挙げられる。T-Connect[2]では、目的地までのナビゲーションや、車から共有された情報を基にしたメンテナンスの案内、異常検知によるカーセキュリティサービスの提供が行われている。また、プローブ情報システムに関する研究も盛んに行われている[5-15]。まず、プローブ情報システムの実用化に向けた検討に関する研究[5-7]として、プローブ情報システムによる情報提供に向けた高速なマップマッチング手法の提案[5]、プローブ情報の増大に対するデータ通信量の抑制に向けたデータ圧縮方式の提案[6]、プローブ情報システムに必要なデータベースシステムの要件の検討[7]が行われている。次に、プローブ情報の活用に関する研究[8-15]について述べる。文献[8-10]では、車から収集された速度やハンドル角などの車両挙動に関するデータを活用することで、特定の道路で起こりやすい車両挙動の分析[8][9]や、運転支援に向けたドライバの運転意図の抽出[10]が行われている。文献[11-13]では、車から収集された位置情報を活用することで、交通量の推定が行われている。文献[14,15]では、車載カメラや車載スマートフォンによって収集された動画画像を活用することで、ドライバ間での動画画像による交通状況の共有が行われている。

本研究では、プローブ情報システムの普及を想定し、交通状況を形成する車両から収集した車載センサ情報により車両挙動を推定し、複数の車両の挙動推定の結果による集合知を用いて、交通状況を形成する車群の挙動（時間的変化）の把握を最終目的とする。既存のカーナビゲーションシステムを利用することによって、ドライバは交通状況を事前に把握し、渋滞していない道路を選択し走行できる。しかし、渋滞していない道路を選択した場合においても、快適な運転を妨げられてしまうことがある。例えば右折待ちをしている車列が走行経路上に存在する交通状況が挙げられ、その車列に接近する車両のドライバは、非効率な減速や車線変更を強いられてしまうことがある。こうした渋滞をドライバが回避するには、走行予定経路上の交通状況を形成する車群の挙動を把握し、事前に走行する車線を判断する必要があると考える。

本研究では車載スマートフォンで収集したプローブ情報に SAX(Symbolic Aggregate Approximation)[16]を適用して、時系列センサデータの文字列への変換を行い、自然言語処理を用いることで、運転行動の時系列性を考慮した車両挙動の推定に向けた特徴抽出を行う。車両挙動データとして、車載スマートフォンからセンサデータを収集する。車両挙動の例として、直進、停止、右/左折、車線変更などが挙げられる。本研究においては、車両挙動をペダル操作による加減速や、ハンドル操作による右/左折といったドライバの一連の運転行動の時系列の結果として成り立つと考える。運転行動の時系列性を考慮した車両挙動推定を行うことで、右折待ち車列の存在の有無だけでなく、その形成過程や車列を回避する車両の挙動などに着目することができると考える。その結果、ドライバの意図を考慮することができ、より詳細な車群の挙動を把握する上で有効であるとする。本稿では、交通状況把握のための車両挙動推定に向けて、車両挙動データから特徴抽出を行う。

†1 公立はこだて未来大学システム情報科学研究科
Graduate School of Systems Information Science, Future University Hakodate
†2 公立はこだて未来大学システム情報科学部
School of Systems Information Science, Future University Hakodate
a) g2117055@fun.ac.jp
b) siraisi@fun.ac.jp

著者らの先行研究[17]では、SAXを適用してセンサデータを抽象化した文字列（以下、センサデータ文字列）をランレングス符号化し、ランレングス符号の数値部分を除去した文字列（以下、センサデータ変動文字列）から車両挙動の特徴抽出を行った。センサデータ変動文字列に出現する部分文字列は、車両挙動を構成する運転行動によるセンサデータの変動に対応していると仮定し、「出現頻度」に着目することで、各車両挙動固有の特徴を表す運転行動を部分文字列として抽出した。その結果、右／左折や減速といったセンサデータの変動が大きい運転行動は、特徴的な部分文字列として抽出することができた。一方で、右／左寄せや右／左車線変更といったセンサデータの変動が小さい運転行動は、それらの運転行動間で同様な部分文字列が出現したことで、各車両挙動にのみ出現する運転行動を、特徴的な部分文字列として抽出することができなかった。先行研究[17]における提案手法の課題として、車両挙動を構成する運転行動による「センサデータの変動時間の違い」を考慮できていないことが挙げられる。

そこで、著者らは、車両挙動を構成する運転行動による「センサデータの変動時間の違い」がランレングス符号の数値部分に表現されていると仮定する。各車両挙動にのみ出現する運転行動を特徴的な部分文字列として抽出するために、センサデータ文字列に対するランレングス符号化の結果（以下、ランレングス符号文字列）に着目することで、「出現頻度」だけでなく、センサデータの変動時間にあたる「時間長」を考慮する。車両挙動を構成する運転行動によるセンサデータの変動部分と、直進部分やノイズによるセンサデータの変動部分と比較して、該当する部分文字列の時間長が長い傾向が見られるため、ランレングス符号文字列においても「時間長」の観点で特徴的であると考える。

本稿では、ランレングス符号文字列に出現する各部分文字列に対して、MDL(Minimum Description Length)原理[18]に基づいた記述長 DL (Description Length) による評価を行うことで、部分文字列の「時間長」と「出現頻度」を考慮し、車両挙動の特徴として抽出すべき部分文字列の抽出を行う。将来的には車両挙動の推定に向けて、本稿で抽出した部分文字列に対して、運転行動に要する時間の違いを考慮することで特徴的な部分文字列として抽出する。

2. 関連研究

2.1 車両挙動データからの特徴抽出に関する研究

車両挙動データを対象にして特徴抽出を行う研究として、文献[8]では、車両挙動データに対して、画像解析分野で用いられる手法である BoS(Bag of Systems)を適用することで、特徴的な車両挙動のパターンを抽出している。この手法では、「緩やかな右転回」といったドライバの意図が反映された車両挙動を抽出している。また、道路の一定区間ごとで、

抽出したパターンの分布をクラスタリングにより求めることで、その道路区間で起きやすい車両挙動を推定している。文献[10]では、車両挙動データに対して二重分節解析を行うことで運転行動パターンを抽出し、センサデータ上の運転行動を意味のあるパターンに分割している。文献[9]では、車両挙動データの集合を「地理情報テンソル」と呼ばれる多次元ベクトルに変換し、隠れマルコフモデルを用いることで、車両挙動のパターン抽出を行っている。地理情報テンソルとは、車両走行データ（時間、緯度経度、センサによる計測値）を基にして、trip（特定の車両による1つの目的を持った出発地点から到着地点までの移動）、zone（道路ごとの車両走行の特徴を抽出するために、道路ごとに設定された小さな区域）、object（各種センサによる計測値）の3つの要素により表現される。MDL原理に基づいて、車両挙動データを表現するモデルの各種パラメータ値の最適解を求めることで、車両挙動のパターンを自動抽出している。

文献[8][10]の手法では、教師なし学習を用いた手法であるため、抽出されたパターンの意味の理解が課題とされている。文献[9]の手法においても、同様な課題があると考えられ、抽出されたパターンが、どのような運転行動に対応しているかという情報が得られなければ交通状況把握への応用は難しいと考える。

2.2 抽象化したデータに対する特徴抽出に関する研究

時系列センサデータに、SAX(Symbolic Aggregate Approximation)を適用することで、文字列に抽象化したデータから特徴抽出を行う研究[19-21]がある。時系列センサデータにSAXを適用して抽象化し、類似した時系列センサデータを同様な文字列として扱うことで、特徴的なパターンを抽出している。各文字が割り当てられる時系列センサデータの値域から、どのような時系列センサデータに対応しているかという情報が得られるため、抽出された特徴的なパターンの傾向を理解しやすいと考える。

文献[19]では、車載センサ情報に対して、SAXを適用してセンサデータの文字列への抽象化を行って車両挙動を文字列のパターンとして扱い、閾値を用いた文字列照合を行うことで急な加減速、右／左折、Uターン、右／左車線変更といった運転行動、車両挙動の検出を行っている。

文献[20][21]では、時系列センサデータに対して、SAXを適用して文字列への抽象化を行い、文字列中に存在する各部分文字列に対して、MDL原理に基づいた評価を行うことで、頻出パターンを抽出している。文献[21]では、モーションキャプチャシステムにより取得した人物動作情報から、動作認識のための頻出動作パターンを抽出している。高次元データである人物動作情報に対して、PFA(Principal Feature Analysis)を適用して低次元化を行い、SAXを適用することで文字列への抽象化を行っている。文字列中に存在する部分文字列を頻出動作パターンの候補として、各部分

文字列に対して MDL 原理に基づいた評価を行うことで、時間長を考慮した頻出動作パターンを、ボトムアップ的なアプローチで自動抽出している。

文献[19-21]でセンサデータに対して適用されている SAX は、メモリ消費や実行時間の削減、センサデータに含まれるノイズの除去、センサデータの容量の削減といった利点が挙げられており、今後コネクテッドカーの普及によりプローブ情報システムで扱われるデータ量が膨大になることを踏まえると、有用であると考えられる。文献[19]では、文字列として抽出されたパターンが、どのような運転行動に対応しているかといった情報が明確であることから、交通状況把握への応用が容易であると考えられる。しかし、車両挙動を、ペダル操作による加減速やハンドル操作による右／左折といった運転行動の時系列として分析していない。そのため推定した車両挙動にドライバーの意図が含まれておらず、交通状況の形成過程や車列を回避する車両の挙動といった本研究が目指す交通状況把握に必要な情報を得るためには、これらの手法を適用することは不十分であると考えられる。また、文献[20][21]では、「時間長」を考慮した特徴抽出が可能であると考えられるが、同様な波形の繰り返しで構成されている単一のセンサデータを対象として評価を行っている。したがって、本研究が対象とする車両挙動データに対して、これらの提案手法を適用することは不十分であると考えられる。

3. 提案手法

3.1 研究目的

本研究では、ドライバーの快適な運転支援に向けて交通状況を形成する車群の挙動（時間的変化）を把握することを最終目的とする。車両挙動は、ペダル操作による加減速や、ハンドル操作による右／左折といったドライバーの一連の時系列的な運転行動の結果として成り立つと考える。したがって、車両挙動をペダル操作やハンドル操作といったドライバーの意図が反映された運転行動の単位で抽出し、車両挙動の推定を行う。そして、推定結果の集合知により、交通状況を形成する車群の挙動の把握を行う。本稿では、交通状況把握のための車両挙動推定の要素技術として、車両挙動データに、SAXの適用による文字列への抽象化とランレングス符号化を行い、MDL(Minimum Description Length)原理に基づいて「時間長」と「出現頻度」に着目した特徴抽出手法の検討を行う。

3.2 交通状況に関する定義

本研究では、交通状況を、複数の車両により形成される車群の挙動と定義する。道路上には、直進する車両、右／左折する車両といった様々な挙動を行う車両が存在している。図 1 に渋滞に関する交通状況の様子を示す。例えば図

1 では、右折車両が多いが、道路上に右折レーンがない状況で発生し得る状況であり、対向車線の車両の通過待ちのために右折待ちの車列が形成され、その車列を回避する車両も発生する。同様の交通状況は、交差点や左折車両が多い場合でも発生し得る。本研究では、渋滞に関する交通状況は、渋滞を形成する車両群と、それを回避する車両群によって形成されていると考える。また、本研究では、車両挙動をドライバーの運転行動の時系列と定義する。例えば、「右折」の挙動は長期的な動きとして捉えると、図 2 のように車道の右側へ車を寄せ（以下、右寄せ）、減速し、右折するという一連の短期的な動きの時系列として成り立つ。本研究では、車両挙動の部分的な箇所（図 2 における右に寄せる、減速する、右折する）を運転行動と定義する。ドライバーが右寄せを始めた地点には個人差があり、ドライバーの意図が反映されていると考えられるため、右折を運転行動の時系列として扱うことで、より詳細な車群の挙動（時間的変化）の把握が可能であると考えられる。

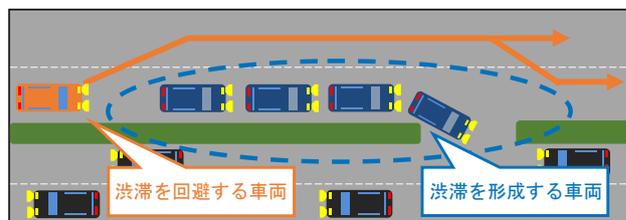


図 1 渋滞に関する交通状況

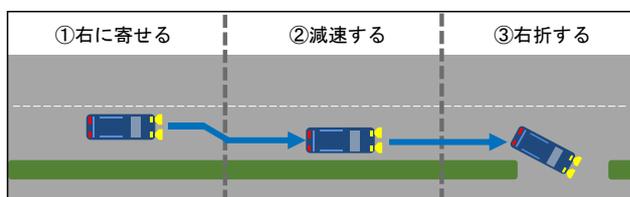


図 2 「右折」挙動の運転行動の時系列

予備実験として、運転行動の時系列性が成り立つかどうかを確認するために、車載スマートフォンを用いてセンサデータを収集した。図 3 において、前後方向の車両の動きを捉える加速度センサの値には、ペダル操作の結果として「減速」の影響が表れており、横方向の車両の動きを捉える角速度センサの値には、ハンドル操作の結果として「右寄せ」や「右折」の影響が表れている。また、図 3 より、図 2 における「右折」挙動を構成する各運転行動が、センサデータ上で時系列的な変化として現れていることがわかる。したがって、本研究では交通状況を形成する車群の挙動把握に向けて、車載スマートフォンから、車両の挙動に関する情報として、センサデータ（加速度、角速度、位置情報、タイムスタンプ）を収集し、車両挙動分析を行う。

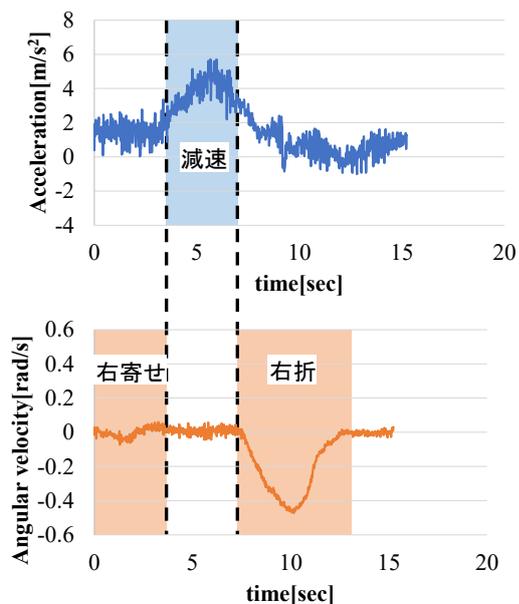


図3 センサデータにおける運転行動の時系列

3.3 研究課題とアプローチ

本稿では、交通状況把握のための車両挙動推定の要素技術の確立に向けて、ランレングス符号文字列に存在する各部分文字列に対して「時間長」を考慮した評価を行い、特徴抽出に向けた検討を行う。研究課題を以下に示す。

課題1. 導入コストの削減

課題2. 運転行動の時系列性の考慮

課題3. 「時間長」に着目した特徴的な運転行動の抽出

課題1について、本研究ではプローブ情報システムを想定しており、より多くのドライバーからプローブ情報を収集するためには、導入コストが低いことが望ましい。そこで、課題1に対するアプローチとして、現在普及しているスマートフォンを利用してプローブ情報（車載センサ情報）を収集する。スマートフォンには、加速度センサ、角速度センサなど多くのセンサが搭載されているため、車両挙動推定に必要な情報を収集できると考える。

課題2について、運転行動の時系列性を考慮するために、収集した車載センサ情報から、運転行動に対応する部分を抽出する必要がある。そこで課題2に対するアプローチとして、車両挙動に対応する時系列センサデータを文字列へ変換し、その変換された文字列からドライバーの運転行動に関する部分文字列を抽出する。具体的には、収集したセンサデータにSAX(Symbolic Aggregate Approximation)[16]を適用してセンサデータを抽象化した文字列（センサデータ文字列）へ変換する。そして、センサデータ文字列に対してランレングス符号化を行うことで、文字部分と数値部分から成るランレングス符号文字列へ変換し、部分文字列を抽出する自然言語処理技術である N-gram を適用する。SAX

により抽象化された車両挙動に関する文字列から、車両挙動を構成する運転行動を、N-gram を適用して部分文字列として抽出することで、車両挙動を運転行動の時系列として表現できると考える。

課題3について、著者らの先行研究[17]の課題である、車両挙動を構成する運転行動による「センサデータの変動時間の違い」の考慮に向けて、「時間長」と「出現頻度」の観点から各車両挙動の特徴として抽出すべき部分文字列の抽出を行う。新たに「時間長」に着目することで、車両挙動を構成する運転行動による「センサデータの変動時間の違い」の考慮を目指す。そこで課題3に対するアプローチとして、課題2に対するアプローチによって得られたランレングス符号文字列に出現する各部分文字列に対して、MDL原理に基づいた記述長 DL による評価を行う。先行研究[17]では、ランレングス符号の数値部分を除去した文字列を用いることで、ランレングス符号の文字部分に着目し、「出現頻度」に着目した特徴的な部分文字列の抽出を行った。本稿では、ランレングス符号の文字部分だけでなく、ランレングス符号文字列（ランレングス符号全体）を用いることで、「時間長」と「出現頻度」の観点から各車両挙動の特徴として抽出すべき部分文字列の抽出を行う。将来的には、本稿で抽出した部分文字列に対して、運転行動に要する時間の違いを考慮することで、特徴的な部分文字列として抽出することを目指す。

3.4 提案手法の流れ

提案手法は、データ収集フェーズと車両挙動分析フェーズから構成される。データ収集フェーズでは、車載スマートフォンを用いて車両走行時のセンサデータの収集を行う。詳細は、3.5節で述べる。車両挙動分析フェーズでは、交通状況把握の要素技術として、車両挙動の推定に向けた車両挙動に関する特徴抽出を行う。車両挙動に関する特徴抽出の手順として、まず SAX を、収集したセンサデータに対して適用し、文字列への抽象化を行う。詳細は、3.6節で述べる。次に、車両挙動に関する特徴的な部分文字列の抽出にむけた前処理として、ランレングス符号化と N-gram による部分文字列の抽出を行う。詳細は、3.7節で述べる。そして、各部分文字列に対して MDL 原理に基づいた記述長 DL による評価を行うことで、「時間長」と「出現頻度」の観点から、各車両挙動の特徴として抽出すべき部分文字列の抽出を行う。詳細は、3.8節で述べる。

先行研究[17]における提案手法との違いとして、3.7節以降で述べる手順により、ランレングス符号文字列を用いることで、「時間長」と「出現頻度」を考慮した特徴抽出を行っている点が挙げられる。データ収集方法（3.5節）、収集したセンサデータに対する SAX の適用方法（3.6節）については、先行研究[17]と同様である。

3.5 車両挙動に関するデータ収集

本研究では、スマートフォンを車両のダッシュボード上に設置し、車両走行時のセンサデータを収集する。スマートフォンに搭載された加速度センサ、角速度センサからセンサデータを取得する。収集するセンサデータは、3軸加速度センサ値、3軸角速度センサ値、緯度、経度であり、タイムスタンプと共に記録する。スマートフォンの加速度センサ、角速度センサのX軸が車両の左右方向、Y軸が鉛直方向、Z軸が進行方向となるようにスマートフォンを設置する(図4)。

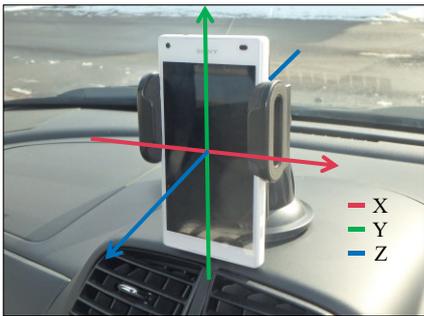


図4 スマートフォンの設置方法とセンサの各軸の向き

3.6 SAXによるデータの抽象化

SAXとは、時系列データを文字列に変換することで、データの抽象化を行う手法である[16]。SAXの手順を以下に示す。

- (I) 時系列データの正規化を行い、時間軸に対して等間隔の区間に分割する。
- (II) 正規分布に従って、正規分布の各面積が等しくなるような境界を定め、区分ごとに文字を割り振る。
- (III) 区分ごとに時系列データの平均値を算出する。
- (IV) 算出した平均値を区分に基づき、文字に変換する。

SAXのパラメータとして、変換率 C と文字の種類 W を用いた。変換後の文字列の長さ L_{after} と、変換する前の時系列データの長さ L_{before} により、 $C=L_{after}/L_{before}$ として計算する。例えば、 $C=1/5$ は、連続する5つのセンサデータで平均値を算出し、1文字に変換を行うことを意味する。また、 $W=4$ の時は、 $\{a,b,c,d\}$ の4種類の文字を利用して、センサデータを文字列に変換することを意味する。SAXは、 W を増減させることで、時系列データの抽象度を変化させることができる。 W が小さいほど、センサデータの抽象度は高くなり、 W が大きいほど、センサデータの抽象度は低くなる。図5にSAXによる抽象化の例($W=4$)を示す。

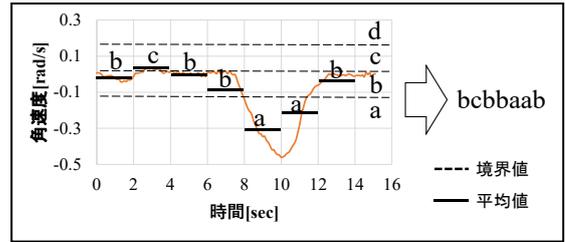


図5 SAXによる抽象化の例

3.7 文字列データに対する前処理

3.6節で述べた手順でSAXを適用してセンサデータを抽象化した文字列(以下、センサデータ文字列)から、「時間長」と「出現頻度」に着目することで、車両挙動を構成する特徴的な部分文字列を抽出する。特徴的な部分文字列の抽出にむけて、以下の手順でセンサデータ文字列に対して前処理を行う。

- Step(I) センサデータのランレングス符号化
- Step(II) N-gramによる部分文字列の抽出

3.7.1 センサデータのランレングス符号化

図6に、センサデータ文字列のランレングス符号化による変換(Step(I))と部分文字列の抽出(Step(II))の様子を示す。

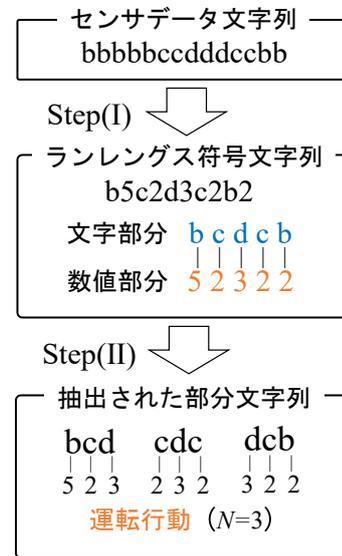


図6 ランレングス符号化と部分文字列の抽出

センサデータ文字列に対して、ランレングス符号化を行い、文字部分と数値部分から成るランレングス符号文字列へ変換する(図6におけるStep(I))。ランレングス符号文字列の数値部分は、各車両挙動における運転行動によるセンサデータの変動時間に対応していると考えられる。よって本稿では、ランレングス符号文字列に着目し、「出現頻度」

だけでなく、「時間長」も考慮することで、車両挙動の特徴として抽出すべき部分文字列の抽出を目指す。

3.7.2 N-gram による部分文字列の抽出

車両挙動を運転行動の時系列として表現するために、ランレングス符号文字列に対して N-gram を適用することで、車両挙動における各運転行動を部分文字列として抽出する(図 6 における Step(II)). N-gram とは、N 文字で構成される部分文字列を元の文字列から抽出する処理である。N-gram のパラメータとして、抽出する部分文字列の長さ N がある。N-gram によって抽出した部分文字列は、運転行動にあたる局所的なセンサデータと対応していると考えられる。そのため、各車両挙動のセンサデータから、各運転行動を抽出できると考える。ここで、運転行動によるセンサデータの変動時間は、ランレングス符号文字列において、運転行動に対応する部分文字列の時間長の総和と定義する。例として、図 6 における Step(II)で抽出した部分文字列 bcd の時間長は、10 (= 5 + 2 + 3) となる。抽出した各部分文字列に対して、3.8 節で定義する MDL(Minimum Description Length)原理に基づいた記述長 DL による評価を行う。

3.8 時間長に着目した文字列データからの特徴抽出

3.8.1 MDL(Minimum Description Length)原理

MDL 原理とは、Rissanen により提案された、「データ圧縮問題だけではなく、多くの工学の分野で適用が可能な原理」[18]である。データ圧縮においては、データを圧縮する際に用いるモデルの記述長と、そのモデルにより圧縮されたデータの記述長の和から成り立つ符号語長を最小にするモデルを選択するという原理である。文献[20][21]では、人間の動作データを SAX で変換した文字列から、MDL 原理に基づいた記述長 DL (Description Length) が最小となる部分文字列を、特徴的な部分文字列として抽出している。記述長 DL が最小となる部分文字列は、その部分文字列を 1 文字の特徴的な文字 (以下、ラベル) に置換した際に、元の文字列を最も圧縮できる部分文字列を示す。記述長 DL の式(1)を、任意の文字列を s 、モデルを h とした場合、データの記述長を式(2)、モデルの記述長を式(3)で表現し、以下のように定義した。

$$DL = DL1(s|h) + DL2(h) \quad (1)$$

$$DL1(s|h) = \sum_i^m \sum_j -l_{ij} \log_2 \frac{l_{ij}}{t_i} \quad (2)$$

$$DL2(h) = \sum_i^m \log_2 t_i + m \log \sum_i^m t_i \quad (3)$$

モデル h は、任意の文字列 s をある部分文字列で圧縮した際の文字列を分割する位置を表現している。本稿では、モデルにより分割された各部分文字列を、セグメントと定義する。 m はセグメント数、 l_{ij} は i 番目のセグメント内で、 j 番目に出現する文字のセグメント内での時間長の総和、 t_i は i 番目のセグメントの時間長の総和を表す。図 7 に、例として部分文字列 bab で分割した場合の文字列の様子を示す。

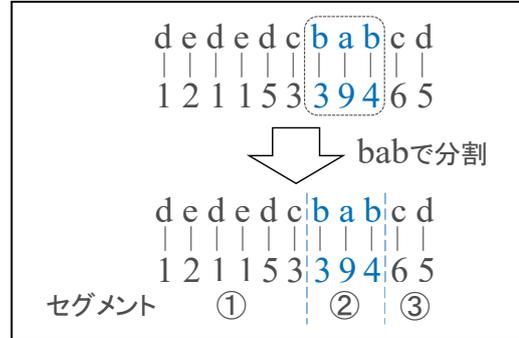


図 7 文字列の分割の様子

図 7 の場合には、 $h = \{6,9\}$ (bab の始点が 6、終点が 9 であることを指す)、 $m = 3$ 、 $l_{11} = 7$ (1 番目のセグメントにおける d の時間長の総和)、 $l_{12} = 3$ (1 番目のセグメントにおける e の時間長の総和)、 $l_{13} = 3$ (1 番目のセグメントにおける c の時間長の総和)、 $l_{21} = 7$ 、 $l_{22} = 9$ 、 $l_{31} = 6$ 、 $l_{32} = 5$ 、 $t_1 = 13$ (1 番目のセグメントの時間長の総和)、 $t_2 = 16$ 、 $t_3 = 11$ となる。任意の部分文字列に対する記述長 DL は、任意の文字列の時間長の総和に対する「時間長」と「出現頻度」の影響を受ける。任意の部分文字列の「出現頻度」が高いほど、セグメント数が大きくなり、ある部分文字列の「時間長」が長いほど、セグメント内での時間長の総和が大きくなる傾向がある。よって、MDL 原理に基づいた記述長 DL は、「時間長」と「出現頻度」を考慮した各車両挙動の特徴として抽出すべき部分文字列の抽出に適用できると考える。

本研究では、3.7.2 項で抽出した全ての部分文字列に対して、記述長 DL を算出する。そして、各部分文字列の記述用 DL を比較することで、「時間長」と「出現頻度」の観点から各車両挙動の特徴として抽出すべき部分文字列の抽出を行う。

3.8.2 MDL 原理に基づいた記述長 DL による特徴抽出

MDL 原理に基づいた記述長 DL を算出し、「時間長」と「出現頻度」の観点から部分文字列の順位付けを行うことで、各車両挙動の特徴として抽出すべき部分文字列の抽出を行う。MDL 原理に基づいた記述長 DL による特徴抽出手法の流れを以下に示す。また、特徴抽出手法の適用したイメージを図 8 に示す。

MDL 原理に基づいた記述長 DL による特徴抽出手法

- 手順1. 出現頻度が 8 割以上の部分文字列を，ランレングス符号文字列から候補として抽出
- 手順2. 候補中の各部分文字列に対して記述長 DL を算出
- 手順3. 記述長 DL が最小となる部分文字列を特徴として抽出し，ラベルに変換
- 手順4. 変換後の文字列から N-gram で部分文字列を抽出し，抽出結果の中でも，ラベルを含まない部分文字列を候補として抽出
- 手順5. 候補中の各部分文字列に手順 1~3 を適用
- 手順6. ランレングス符号文字列中に，文字列長が 2 以上の部分文字列が出現しなくなるまで，手順 4~5 を繰り返す



図 8 MDL 原理による特徴抽出手法のイメージ

手順 1 で，著者らの先行研究[17]で定義した特徴部分文字列と準特徴部分文字列を「出現頻度」の観点から特徴的な部分文字列として抽出し，MDL 原理による評価候補とする。以下の条件 C1，条件 C2 を満たした部分文字列を特徴部分文字列として抽出する。また，特徴部分文字列の条件 C1 を満たすが，条件 C2 を満たさない部分文字列を準特徴部分文字列とする。本稿では，特徴抽出部分文字列または準特徴部分文字列に該当する部分文字列を，特徴的な部分文字列と定義する。以下の条件を満たす部分文字列は，各車両挙動固有の特徴を表す部分文字列であると考えられる。

特徴部分文字列の条件

- 条件 C1** 各車両挙動のセンサデータの 8 割以上で出現する。
- 条件 C2** 他の車両挙動のセンサデータでは，出現しない。

条件 C1 において，各車両挙動のセンサデータの 8 割以上で出現する部分文字列を特徴的な部分文字列とした理由として，同様な車両挙動に関するセンサデータの値が，試行ごとでわずかに異なる傾向にあったことが挙げられる。センサデータの値が異なると，SAX を適用した際に変換される文字も異なってしまう場合がある。センサデータの値が異なる要因として，同一ドライバでも，同じ車両挙動の試行ごとで差が生じてしまうことが考えられる。また，プローブ情報システムを想定すると，ドライバや車種の違い，走行する環境の違いによって，同様な車両挙動に関するセンサデータの値が異なってしまうことが考えられる。

そこで，「出現頻度」の観点で各車両挙動固有の特徴を抽出する際に基準値を設定する。この基準値以上の頻度で出現する部分文字列を対象とすることで，様々な要因によるセンサデータの違いを許容する。本稿では，「8 割以上」と設定しているが，今後の実験を通して，適正值を検討する予定である。

手順 2 で，評価候補として抽出された各部分文字列に対して，3.8.1 項で定義した記述長 DL を算出する。

手順 3 で，算出した記述長 DL が最小な部分文字列を「時間長」と「出現頻度」の観点から最も特徴的な部分文字列として抽出し，ラベルに変換する。図 8 では，部分文字列 bc に対する記述長 DL が最小となった場合を示しており，部分文字列 bc を最も特徴的な部分文字列として抽出する。そして部分文字列 bc をラベル P₁ に変換している。

手順 4 で，ラベルを含んだ変換後の文字列に対して N-gram で部分文字列を抽出し，抽出結果の中でもラベルを含まない部分文字列を MDL 原理に基づいた記述長 DL による評価候補とする。

手順 5 で，評価候補として抽出された各部分文字列に対して，手順 1~3 を適用し，2 番目に特徴的な部分文字列を抽出する。図 8 では，部分文字列 df に対する記述長 DL が最小になった場合を示しており，部分文字列 df を 2 番目に特徴的な部分文字列として抽出する。そして，部分文字列 df をラベル P₂ に変換している。そして手順 6 として，文字列中に，長さが 2 以上の部分文字列が出現しなくなるまで，手順 4~5 を繰り返し，3 番目以降に特徴的な部分文字列の抽出を行う。その結果として，図 8 中の抽出結果のように，特徴的な部分文字列を抽出することができる。

以上の特徴抽出手法を適用することで，「時間長」と「出現頻度」の観点に基づいて，各車両挙動の特徴として抽出すべき部分文字列の抽出ができるようになる。また，それらをラベルに変換することで，明示的に文字列中に各車両挙動の特徴として抽出すべき部分文字列の出現を表現することができる。これは，将来的に各車両挙動の特徴として抽出すべき部分文字列の特定につながるため，有用であると考えられる。

4. 実験および考察

4.1 センサデータの収集

センサデータ収集時の実験条件を表 1 に示す。

本稿の実験では、車両 1 台 (TOYOTA パッソ) に、図 4 に示すようにスマートフォンを設置し、被験者 1 名 (20 代男性) の運転時のセンサデータを収集した。本研究で対象とする交通状況を構成する代表的な車両挙動として、直進、右折、左折、右車線変更、左車線変更の 5 種類を収集対象とした。本研究では、道路を一定間隔で区切った区間ごとに、交通状況把握を行うことを想定しており、また実環境に近いデータを収集する必要があると考える。そこで、一般道における最高速度である時速 60km/h に達した地点から 100m 走行し、その区間内でペダル操作やハンドル操作を含む各車両挙動を行った際のセンサデータを、各車両挙動 20 回分について収集した。実験環境として、直進、右/左車線変更は直線道路、右/左折は T 字路でセンサデータの収集を行った。道路の凹凸による振動の影響を考慮して、道路の舗装が整備された環境で実験を行ったため、本稿では道路の凹凸によるノイズの影響は考慮しないものとする。

表 1 センサデータ収集時の実験条件

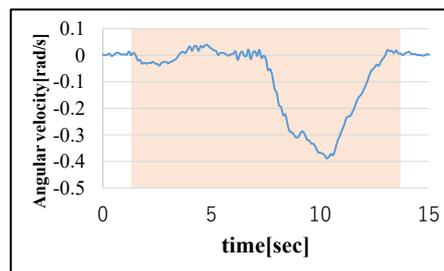
使用センサ	加速度, 角速度, GPS
サンプリングレート (加速度, 角速度)	50Hz
サンプリングレート (GPS)	1Hz
対象とする車両挙動	直進, 右折, 左折 右車線変更, 左車線変更
各車両挙動の収集回数	20 回
収集時の速度	60km/h
収集時の走行距離	100m

4.2 提案手法の適用に向けた前処理

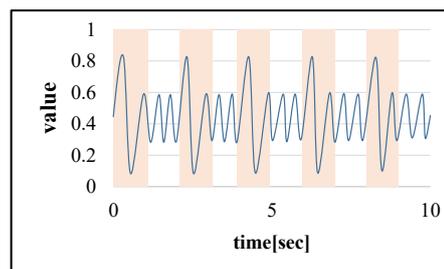
3.8 節で述べた提案手法を適用する際に、3.5 節で収集したセンサデータでは各車両挙動が 1 度しか出現しないことが問題となる。図 9 に、対象とするセンサデータの比較を示す。文献[20][21]では、図 9(b)のような繰り返しのあるセンサデータを対象に特徴抽出を行っている。

図 9 に示すように、本研究では収集したセンサデータには各車両挙動が 1 度しか出現しないため、MDL 原理に基づいた記述長 DL による各車両挙動の特徴として抽出すべき部分文字列の抽出は難しいと考える。そこで、収集した各車両挙動 20 回分のセンサデータを 1 つの時系列センサデータとして扱う。繰り返しのある車両挙動データとして扱うことで、文献[20][21]で対象とするセンサデータと同様に、繰り返しのあるセンサデータとして扱うことができる。本稿では、各車両挙動で収集した 20 回分の同一車両挙動のセ

ンサデータそれぞれに対して、3.8.1 項で定義した記述長 DL の式(1)を算出し、その総和をある部分文字列に対する記述長 DL として採用した。



(a) 本研究の対象センサデータ (右折時の Y 軸角速度)



(b) 文献[20][21]における対象センサデータのイメージ

図 9 対象とするセンサデータの比較

4.3 提案手法の評価

本節では、3.5 節の手順で収集したセンサデータ (Y 軸角速度, Z 軸加速度) に対して提案手法による記述長 DL の算出結果を示す。MDL 原理の性質上、文字列長が 2 以上の部分文字列が圧縮の対象になるため、N=1 の部分文字列は、今回は抽出の対象外とする。表 2 に今回適用した、各種パラメータ値を示す。

表 2 分類実験に用いる各種パラメータ

SAX のパラメータ	
変換率 C	1/5
文字の種類 W	8
N-gram のパラメータ	
部分文字列の長さ N	2,3

4.3.1 左右方向に関する運転行動への適用結果

本項では、右/左折などの左右方向に関する運転行動の影響が見られた Y 軸角速度のセンサデータに関する抽出結果について述べる。本稿では、特に右折時のセンサデータに対する提案手法の適用結果 ($W = 8$) を述べる。3 章で述べた提案手法を用いて特徴抽出を行うが、右折における左右方向に関する運転行動として、「右寄せ」と「右折」をセンサデータ収集時に行っている。よって、3.8.2 項の提案手法における手順 5 までを適用し、「時間長」の観点から最も特徴的な部分文字列と、2 番目に特徴的な部分文字列を

抽出し、「右寄せ」と「右折」に該当する部分文字列が抽出できるかについて検討する。図 10 に手順 1~3 によって算出された各部分文字列に対する記述長 DL, 図 11 に手順 4~5 によって算出された各部分文字列に対する記述長 DL を示す。

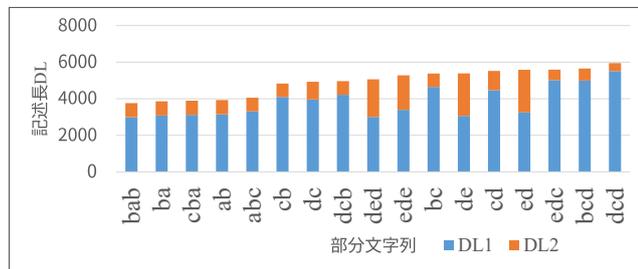


図 10 手順 1~3 で算出された各部分文字列の記述長 DL

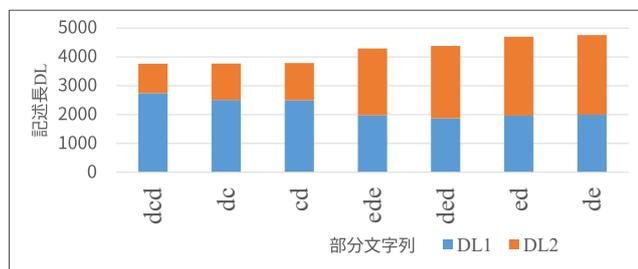


図 11 手順 4~5 で算出された各部分文字列の記述長 DL

図 10 では、部分文字列 bab に対する記述長 DL が最小となっており、部分文字列 bab が最も特徴的な部分文字列であることを示している。また、図 11 では、部分文字列 dcd に対する記述長 DL が最小となっており、部分文字列 dcd が 2 番目に特徴的な部分文字列であることを示している。図 12 に、部分文字列 bab, 部分文字列 dcd に該当するセンサデータの様子を示す。

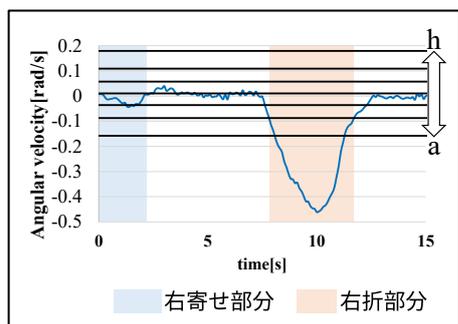


図 12 部分文字列 bab,dcd に該当するセンサデータ

図 12 に示すように、抽出された部分文字列 bab, 部分文字列 dcd とセンサデータの対応関係を分析すると、記述長 DL の算出結果から最も特徴的な部分文字列 (bab) と 2 番目に特徴的な部分文字列 (dcd) は、それぞれセンサデータにおける「右折」と「右寄せ」に該当する部分文字列であ

ることがわかった。

図 13 に、手順 1~3 で算出された Y 軸角速度のランレングス符号文字列における各部分文字列の平均時間長を示す。図 13 における時間長は、各部分文字列のランレングス符号文字列における数値部分の総和の平均値 (各部分文字列の出現回数で平均を算出) を示している。ランレングス符号文字列における数値部分は、同じ文字が連続して出現する回数を表している。例えば数値部分が 5 の場合は、5 文字連続で同じ文字が出現していることを意味する。本稿では、50Hz で収集したセンサデータを $C = 1/5$ で文字に変換しているため、1 文字あたり 0.1 秒分のセンサデータに該当する。

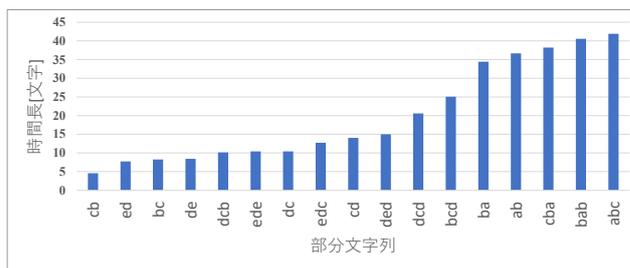


図 13 手順 1~3 で算出された各部分文字列の平均時間長

図 13 から部分文字列 bab の平均時間長が、右折時の Y 軸角速度のランレングス符号文字列における部分文字列の中でも長い傾向にあることがわかる。これは、部分文字列 bab に対する記述長 DL が最も小さくなった一因であると考えられる。部分文字列 bab, dcd は、「時間長」と「出現頻度」の観点から特徴的な部分文字列であり、各車両挙動における運転行動を表しているため、車両挙動推定に有効な特徴であると考えられる。

先行研究[17]において、「出現頻度」の観点から抽出を行った際にも、「右折」に該当する bab は、特徴部分文字列として抽出することができていた。しかし、センサデータの変動が小さい「右寄せ」に該当する部分文字列を、特徴部分文字列として抽出することができなかった。今回の実験結果では、記述長 DL を用いた提案手法により複数の部分文字列の中で、「右寄せ」に該当する dcd を、2 番目に特徴的な部分文字列として抽出することができた。したがって、「時間長」の観点を考慮することの有効性が示唆された。

4.3.2 前後方向に関する運転行動への適用結果

本項では、加減速などの前後方向に関する運転行動の影響が見られた Z 軸加速度のセンサデータに関する抽出結果について述べる。本稿では、特に右折時のセンサデータに対する提案手法の適用結果 ($W = 8$) を述べる。3 章で述べた提案手法を用いて特徴抽出を行うが、右/左折時における前後方向に関する運転行動として、「減速」をセンサデータ収集時に行っている。よって、3.8.2 項の提案手法にお

ける手順3までを適用し、「時間長」の観点から最も特徴的な部分文字列を抽出し、「減速」に該当する部分文字列が抽出されるか検討する。図14に手順1~3によって算出された各部分文字列に対する記述長DLを示す。

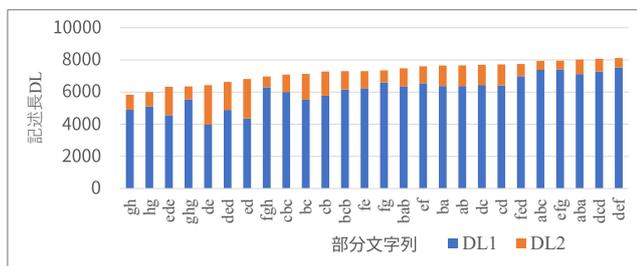


図14 手順1~3で算出された各部分文字列の記述長DL

図14から、部分文字列ghに対する記述長DLが最小となっており、部分文字列ghが最も特徴的な部分文字列であることを示している。図15に、部分文字列ghに該当するセンサデータの様子を示す。

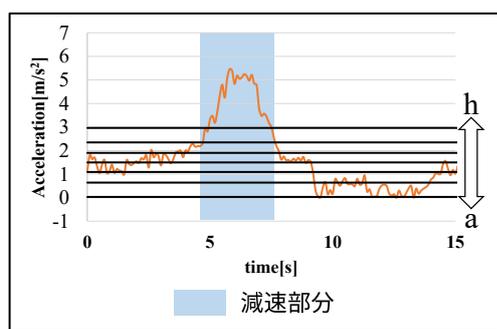


図15 部分文字列ghに該当するセンサデータ

図15に示すように、抽出された部分文字列ghとセンサデータの対応関係を分析すると、抽出された最も特徴的な部分文字列は、センサデータにおける「減速」に該当する部分文字列であることがわかった。

図16に、手順1~3で算出されたZ軸加速度のランレングス符号文字列における各部分文字列の平均時間長を示す。

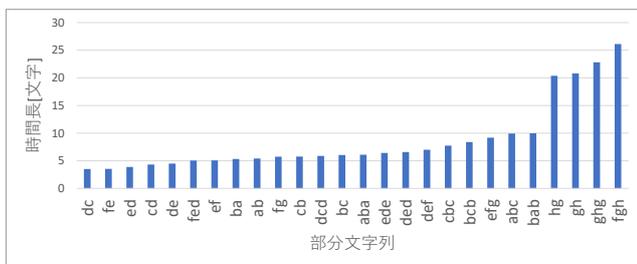


図16 手順1~3で算出された各部分文字列の平均時間長

図16から部分文字列ghの平均時間長が、右折時のZ軸加速度のランレングス符号文字列における部分文字列の中

でも長い傾向にあることがわかる。このことは、部分文字列ghに対する記述長DLが最も小さくなった一因であると考える。部分文字列ghは、「時間長」と「出現頻度」の観点から特徴的な部分文字列であり、各車両挙動における運転行動を表しているため、車両挙動推定に有効な特徴であると考えている。よって、記述長DLを用いた提案手法により、車両挙動の特徴的な運転行動に該当する部分文字列を抽出できることが示唆された。

5. まとめ

本稿では、交通状況を形成する車群の挙動(時間的変化)の把握に関する要素技術として、車両挙動の推定に向けた特徴抽出手法の提案を行った。車載スマートフォンから収集したセンサデータにSAX(Symbolic Aggregate Approximation)を用いることで文字列へ変換し、N-gramによる部分文字列としての運転行動の抽出、そしてMDL(Minimum Description Length)原理に基づく「時間長」と「出現頻度」を考慮した特徴抽出を行った。実験として、提案手法を用いて、右折時のセンサデータ(Y軸角速度、Z軸加速度)から車両挙動に関する特徴抽出を行った。その結果、左右方向の移動を表すY軸角速度からは、「右寄せ」、「右折」に該当する部分文字列、前後方向の移動を表すZ軸加速度からは、「減速」に該当する部分文字列を、それぞれ「時間長」と「出現頻度」の観点から各車両挙動の特徴として抽出すべき部分文字列として抽出することができた。

今後の課題として、各車両挙動を構成する全ての運転行動を特徴的な部分文字列として抽出することが挙げられる。記述長DLの算出結果から、各車両挙動の特徴として抽出すべき部分文字列を特定し、異なる車両挙動において出現する同様な部分文字列を区別することで特徴的な部分文字列として扱うことができると考える。また、車両挙動の推定に向けて、複数のセンサデータストリームにおける運転行動の時系列性を、部分文字列間の関係として表現する必要がある。そのために、今回抽出した部分文字列の、時系列的な出現の順序関係を利用する予定である。

謝辞 本研究はJSPS 科研費JP17K00128の助成を受けたものである。

参考文献

- [1] 植原啓介:プローブ情報システム:車載センサを活用した環境情報の取得, 情報処理, Vol.51, No.9, pp.1144-1149(2010).
- [2] トヨタ自動車株式会社:トヨタ T-Connect, 入手先 <<https://toyota.jp/tconnectservice/>>(参照 2018-7-12).
- [3] 本田技研工業株式会社: Honda internavi, 入手先 <<http://www.honda.co.jp/internavi/>>(参照 2018-7-12).
- [4] 日産自動車株式会社:日産: NissanConnect サービス/サービス概要, 入手先 <http://www.nissan.co.jp/OPTIONAL-PARTS/NAVI_AUDIO/NI

- SSAN_CONNECT/index.html>(参照 2018-7-12).
- [5] 濱岡達哉, 水将人, 横田孝義:プローブカーの位置情報の高速かつ大域的評価によるマップマッチング手法:動的計画法とリンク接続判定の高速化を用いたマップマッチング手法, 電子情報通信学会技術研究報告, Vol.117, No.203, pp.17-22(2017).
 - [6] 加藤翼, 柿澤浩仁, 清原良三:テレマティクスサービスに向けた端末上でのプローブデータ生成方式, 情報処理学会研究報告, Vol.2014-ITS-56, No.4, pp.1-6(2014).
 - [7] 疋田敏朗, 堀口賢司, 栗原慶典ほか:車両センサデータを蓄積・活用するためのデータベースシステムの提案, 情報処理学会研究報告, Vol.2013-DBS-158, No.30, pp.1-7(2013).
 - [8] 柏倉俊樹, 三宮千尋, 田中雄介ほか: Bag of Systems を用いた運転行動解析, 情報科学技術フォーラム講演論文集, Vol.14, No.1, pp.159-162(2015).
 - [9] 本田崇人, 松原靖子, 根山亮ほか:車両走行センサデータからの自動パターン検出, 情報処理学会論文誌, データベース, Vol.9, No.3, pp.1-13(2016).
 - [10] 谷口忠大, 山下元気, 長坂翔吾ほか:運転時系列のベイズ二重分節解析によるチャンク抽出, 計測自動制御学会論文集, Vol.49, No.11, pp.1047-1056(2013).
 - [11] 和田健太郎, 小林桂子, 桑原雅夫:プローブ車両軌跡データのみを用いた交通の量的把握, 生産研究, Vol.67, No.2, pp.143-147(2015).
 - [12] 三輪富生, 山本俊行, 竹下知範ほか:プローブカーの速度情報を用いた動的 OD 交通量の推定可能性に関する研究, 土木学会論文集 D, Vol.63, No.2, pp.252-265(2008).
 - [13] 丹下智之, 廣森聡仁, 梅津高朗ほか:一部の車両軌跡情報および信号パラメータを用いた OD 交通量の推定手法, 情報処理学会研究報告, Vol.2016-ITS-64, No.5, pp.1-8(2016).
 - [14] 伊藤亮輔, 石原進:車載カメラ撮影映像提供システムにおける位置指定要求に対する映像選択方法, 情報処理学会論文誌, Vol.58, No.1, pp.79-88(2017).
 - [15] 玉井森彦, 尾上佳久, 安本慶一ほか:画像処理に基づいた効率のよい渋滞動画の収集・共有方式, 情報処理学会研究報告, Vol.2013-MBL-65, No.36, pp.1-8(2013).
 - [16] Lin, J., Keogh, E., Lonardi S., et al.:A Symbolic Representation of Time Series, with Implications for Streaming Algorithms, In SIGMOD Workshop, pp.2-11(2003).
 - [17] 横山達也, 白石陽:交通状況把握のための運転行動の時系列性を考慮した車両挙動分析手法, 情報処理学会, マルチメディア, 分散協調とモバイルシンポジウム 2017 論文集, Vol.2017, pp.73-82(2017).
 - [18] 下平英寿, 伊藤秀一, 久保川達也ほか:モデル選択:予測・検定・推定の交差点, pp.101-138, 岩波書店(2004).
 - [19] Chaovalit, P., Saiprasert, C. and Pholprasit, T.:A Method for Driving Event Detection Using SAX with Resource Usage Exploration on Smartphone Platform, EURASIP Journal on Wireless Communications and Networking, Vol.2014, pp.1-11(2014).
 - [20] Yoshiki Tanaka, Kazuhisa Iwamoto, Kuniaki Uehara, Discovery of Time-Series Motif from Multi-Dimensional Data Based on MDL Principle, Machine Learning, Vol.58, Issue 2-3, pp.269-300(2005).
 - [21] 荒木裕, 有田大作, 谷口倫一郎:時系列動作情報からの頻出動作パターンの抽出について, 電気学会研究会資料.IIS, 産業システム情報化研究会, Vol.IIS-06, No.19-26, pp39-44(2006).