

フローティングカーデータおよび地理情報を用いた 停車時間予測のための特徴量探索

Feature Analysis for Temporary Stopping Time Prediction Using Floating Car Data and Geographic Information

熊谷 優惟子[†] 木村 光樹[†] 渡邊 和哉[†]
Yuiko Kumagai Mitsuki Kimura Kazuya Watanabe

1 はじめに

近年、IVI の進化に伴い、車両走行中にドライバーが受け取る情報や利用できるサービスは増えてきている。しかし、運転中の情報提示やカーナビゲーションなどの車載情報機器の操作によって、運転への注意は低下する。これは、ドライバーディストラクションとして知られており [1]、安全性の観点から対策が必要となっている。

この対策の一つとして、カーナビゲーションなどの車載情報機器は、国内外で運転中の使用規制がある。日本自動車工業会の画面表示装置に関するガイドライン [2] では、走行中にテレビやビデオ等の運転に関係のない動画や画像の提示を禁止している。また、操作に関しても、タスクが完了するまでの画面総確認時間を 8 秒未満にすることが求められている。ガイドラインに従えば、運転操作に直接関連しない画像提示や、ある程度時間のかかる操作は、停車中に完了させることが望ましい。なぜなら、停車中に運転操作以外の行動が完了しなければ、ドライバーディストラクションを引き起こす可能性があるからである。

そこで、停車時間の長さを予測して提示することで、ドライバーは、安全かつ快適に運転操作以外に取り組みると考えられる。例えば、停車時間が短いと予想される際には、流している音楽を変更したり、目的地の天気を確認するなど簡単な操作を行う。一方、停車時間が長い場合には、飲み物を飲む、ルート上のレストランを検索し経路地に設定するなど、ある程度時間のかかる動作や操作にも取り組むことができる。このように停車時間の長さを予測することで、ドライバーが停車中の時間を安全に有効活用できる。

現在実用化されている、信号情報活用運転支援システム (Traffic Signal Prediction Systems; TSPS) [3] を利用することで、ドライバーは信号が赤から青に変わるまでの残り時間などを知ることが可能である。しかし、TSPS を活用するためには、信号情報を配信するための光ビーコンの整備および TSPS からの情報を受信可能な車載情報機器が必要である。これらの整備や普及にはコストがかかる。

交差点における停止時間に関しては、これまで旅行時間を予測する分野で研究が行われてきた。旅行時間とは、道路を区間に分割し、各区間を通過するのにかかった時間のことである。この旅行時間は、走行時間および停止時間の総和として計算されるため、停止時間の定式化が検討されている [4, 5, 6] が、予

測精度は十分でない。文献 [5, 6] では、渋滞や信号機を制御するパラメータ、信号交差点での車の進行方向 (直進, 右折, 左折) を考慮し、走行時間および停止時間を計算した。そして、この計算で得た旅行時間および試験走行した測定値を比較した。その結果、推定した停止時間と測定値の誤差にはばらつきがあることや [5]、旅行時間は、停止の要因となる赤信号表示時間に影響を受けること [6] がわかった。また、停車の要因は信号待ちや左右折待ちだけでなく、合流、POI (Point of Interests) への進入・退出など様々な事象を含む。このように交差点における停車時間の予測には課題がある。

そこで本研究では、特殊なハードウェアが不要な状態で、様々な要因の停車時間の長さを予測することを目標とする。道路区間ごとに停車時間予測モデルを構築するには多大なコストがかかるため、道路区間共通モデルの構築が必要だと考えられる。本稿では、初期検討として、フローティングカーデータ (FCD) および道路の地理情報を用いて、全道路区間に対して 1 つの共通モデルを構築する。その結果を考察することで、停車時間の長さを予測するのに重要な道路区間共通の特徴量を探索する。

2 分析対象および検討した特徴量

2.1 分析対象

本検討では、停車時間を、カーナビゲーションを操作可能な時間として、「ある速度になった時刻からある速度になった時刻までの時間」と定義した。今回は、0km/h から 5km/h になるまでの時刻とした。

分析には、緯度経度や速度などがレコードとして含まれている FCD および一般社団法人日本デジタル道路地図協会の全国デジタル道路地図データベースを使用した。事前分析で、停車回数が多いと判定された、関東の道路 26 区間 (有料道路を除く) を対象とした。この道路 26 区間の大半は交差点付近であり、道路区間の長さは区間平均 251.9 m、区間最大 532 m、区間最小 51 m である。

まず、FCD から明らかな測定異常値を除去した上で、長期休暇を含む一定期間における停車時間を求めた。抜粋した道路 6 区間の停車時間を図 1 に示す。最終的な対象データ数 (停車回数) は 56,483 データ (区間最大: 4,511、区間最小: 857) である。

2.2 検討した特徴量

FCD と全国デジタル道路地図データベースから考えられる特徴量を作成した。それに加えて、平日と休日に関するフラグを作成した。最後に、全ての特徴量計 31 種類を標準化した。以下に、特徴量を作成した理由と作成方法、略称を述べる。

1) 直前の速度 直前の速度遷移は停車時間の長さに影響がある

[†] 株式会社本田技術研究所, Honda R&D Co., Ltd.

表 1 クラス一覧

特徴量	内容
Class 1	停車時間 40 s 以下
Class 2	停車時間 40 s 以上 70 s 未満
Class 3	停車時間 70 s 以上

と考えられたため、作成した。停車直前の速度として、停車から 1-5 サンプル前の速度 (V1-V5) および停車から 2-5 サンプル前の速度変化量 (VV2-VV5) を FCD から求めた。速度変化量は、停車から N サンプル前の速度より停車から $N-1$ サンプル前の速度を減算した値である。

2) 旅行時間 旅行時間の集計値は、通勤など時間帯により発生する渋滞や信号サイクルの代替値となり、停車時間に影響を与えると考えられるため、作成した。まず、1日を 30 s ごとの時間帯に区切り、それぞれの道路区間ごとに、ある一定期間の同一時間帯で旅行時間を集計した。代表値として、旅行時間の 25 パーセンタイル、中央値、75 パーセンタイル (DT-25, DT-50, DT-75) を算出した。規格化、中心化した値として、それぞれの代表値を道路区間ごとに分散で割った (DT-25V, DT-50V, DT-75V), および道路区間ごとに道路区間の平均値を引いた (DT-25M, DT-50M, DT-75M) 値を算出した。

3) 停車開始地点から道路区間の終端までの距離の調整値 停車位置から交差点の停止線までの距離は停車時間に関係すると考えられるため、作成した。道路区間の終端は、停止線より交差点中心に近いことが多く、停止線の位置の緯度経度は明らかになっていない。そこで、停止線の位置を「各道路区間における停車位置から道路区間の終端までの距離の最頻値」と仮定した。本稿では、FCD より求めた停車位置の緯度経度と全国デジタル道路地図データベースの道路区間終端の緯度経度からヒュベニの公式を用いて距離を算出した。道路区間ごとに、当該距離から、その区間の最頻値を減算した値を使用した (DS)。

4) 道路区間の地理情報 道路の地理的な特徴が停車時間に影響を与えると考えられるために、使用した。全国デジタル道路地図データベースより、リンク長 (LL), リンク種別コード (LT), 車線数コード (LC), ビーコン有無コード (BE), 車道幅員 (RW), 最小車道部幅員 (MRW), 交通規制条件種別コード (TRC), 規制速度コード (TR), ノード種別コード (NT), 接続リンク本数 (CL) を使用した。LL は、道路区間の長さである。LT は、本線 (上下線非分離/分離), 連結路, 交差点内などの道路区間の種別を表す。NT は交差点, 行き止まり点などの道路区間終端の種別コードである。また、CL は道路区間終端に接続する道路の総本数を示す。

5) 平日・休日フラグ 1日あたりの交通量は、平日に比べ休日の方が多くと考えられるため作成した。祝日を含む土日を休日と定義したフラグ (DH) である。

3 モデルの構築と検証方法

本稿では、停車時間の長さを、表 1 に示す 3 クラスに分類する問題として設定した。クラスの閾値は、停車時間の長さの中央値 39 s と 95 パーセンタイルである 72 s を参考に定義した。

表 2 に示すデータセットを作成し、それぞれに対して分類モデルの構築を行った。はじめに、データセットについて説明する。次に、分類モデルおよび検証方法について述べる。

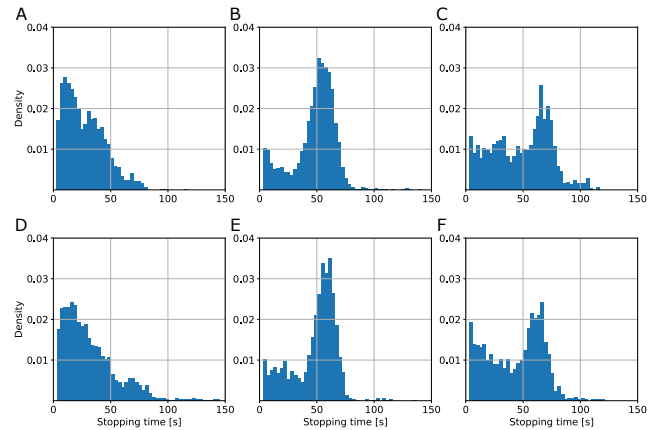


図 1 道路区間ごとの停車時間 (道路 6 区間抜粋). 横軸は停車時間, 縦軸は累積相対度数を示す。

3.1 全データ

道路区間共通モデルの構築するため、全データを用いた。ある一定期間において道路 26 区間で発生する停車全てを含むデータセットである。

3.2 アンダーサンプリングしたデータセット

道路区間のサンプル数を最小の道路区間 (857 サンプル) に合わせるため、アンダーサンプリングを行った。サンプル数の多い道路区間から、ランダムにサンプルを抽出することで、データセットを作成した。

3.3 道路区間ごとの停車時間分布を基に分割したデータセット

図 1 に示すように、道路区間によって停車時間の分布は異なった。図 1A, D のように Class 1 のサンプル数が多い区間、図 1B, E のように Class 2 のサンプル数が多い区間、図 1C, F のように Class 1 と Class 2 のサンプル数が同程度の区間があった。そこで、Class 1 のサンプル数が Class 2 のサンプル数の 1.2 倍以上の区間を Class 1 が多い区間 (Class 1 > Class 2), その一方で、Class 2 のサンプル数が Class 1 のサンプル数の 1.2 倍以上の区間を Class 2 が多い区間 (Class 1 < Class 2), そして、それ以外の区間を Class 1 と Class 2 が同程度の区間 (Class 1 = Class 2) と定義し、データセットを 3 分割した。

3.4 分類モデルおよび検証方法

分類には、教師あり学習の Random Forest (RF) を利用した。RF は機械学習ライブラリの Scikit-learn [7] を使い、パラメータは、決定木の数を 2,000, 決定木の最大深さを 15, 最小ジニ不純度を 0.000005 とした。

モデルの検証として、データセットを訓練データとテストデータにランダムで分割し、学習と評価を 10 回繰り返した。訓練データとテストデータの割合は、7:3 とした。また、モデルの有効性を確認するため、全て最多クラスと判定した際の正解率をベースラインとして算出し、構築したモデルの正解率と比較した。最後に、各モデルにおける特徴量の重要度として、ジニ係数をアンサンブル平均した値を算出し、マクロ平均した。

表 2 モデルの分類結果

データセット名	対象道路区間数	サンプル数	モデルの正解率 [%]			ベースライン [%]
			マクロ平均	最小	最大	
全データ	26	56,483	71.2	70.7	71.7	52.9
アンダーサンプリング	26	22,282	68.3	67.8	68.7	50.9
Class 1 > Class 2	10	23,478	76.5	75.8	77.2	72.7
Class 1 < Class 2	8	18,843	71.5	71.1	72.1	61.1
Class 1 ≐ Class 2	8	14,162	61.7	60.7	62.3	46.3

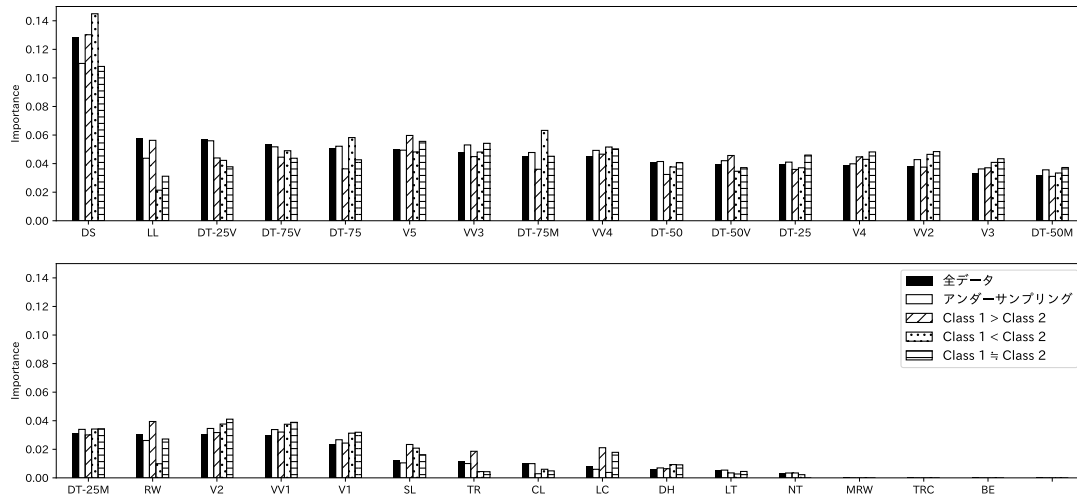


図 2 全 31 種類の特徴量の重要度。全データに対するモデルの特徴量の重要度が大きい順に 2 段に分けて示した。縦軸は重要度の大きさを示す。

4 実験結果および考察

4.1 停車時間の分布

図 1B, C, E, F に示すように 40–70 s に分布のピークがみられた。この各道路区間における一定時間の停車すなわち Class 2 の長い停車は、信号待ちや右折待ちが多く含まれると推測される。一方、Class 1 は、左折待ちや合流などが多く含まれるため、停車時間が短いと考えられる。

4.2 モデルの正解率とサンプル数

構築した各モデルの結果を表 2 に、各モデルの特徴量の重要度をマクロ平均した結果を図 2 に示す。ここで、正解率とサンプル数について相関をみたところ、“全データ”モデルでは、道路区間ごとのサンプル数とモデルの正解率に、 $r = 0.62$ ($p < 0.01$) と、相関が認められた。“アンダーサンプリング”モデルでは、Class 1 および Class 2 のサンプル数とモデルの正解率に、それぞれ $r = 0.45$ ($p < 0.01$)、 $r = -0.55$ ($p < 0.01$) と、相関がみられた。道路区間ごとの停車時間分布を基に分割したモデルにおいても、各クラスのサンプル数と正解率の間で、“Class 1 > Class 2”のモデルは $r = 0.65$ ($p < 0.01$)、“Class 1 < Class 2”のモデルは $r = 0.50$ ($p < 0.01$)、“Class 1 ≐ Class 2”のモデルは $r = 0.69$ ($p < 0.01$) と、相関がみられて

いる。

一方、表 2 の結果より、5 種類のデータセット全てに関して、正解率をマクロ平均した結果および最大、最小正解率に大差はなかった。このことから、検証におけるサンプリングの影響は少ないと考えられる。また、モデルの正解率はベースラインよりも高い。よって、一定の分類性能を有するモデルは構築できていると考えられることから、この結果に基づいて、停車時間を説明する上で重要になる特徴を分析する。

特徴量の重要度を見ると、LL に関しては、アンダーサンプリングモデルよりも全データモデルの方が大きい。LL は、道路区間を通過する際の停車発生確率に影響があると考えられるが、今回は交通量が比較的多い道路区間を選択したため、サンプル数が増加し、リンク長の影響が重要度に現れていると推察される。

一方、全データに対して構築したモデルの道路区間ごとの正解率とサンプル数に相関がみられ、サンプル数が多いほど、正解率が高くなった。特徴量の重要度に関して、アンダーサンプリングしたモデルと比較すると、全データの LL の方が大きくなった。停車時間の分布で分割したデータセット、特に、Class 2 > Class 1 のモデルにおいても、サンプル数に関係する LL の重要度が大きくなった。

今回使用した特徴量では、前述のように、サンプル数と正解率に相関がみられたことから、サンプル数の多い、一部の道路区間に特化したモデルに収束していると考えられ、全道路区間の共通性は表現できていない。汎用的なモデルを構築するためには、さらに特徴量やモデル、パラメータなどの検討が必要である。

4.3 停車時間の予測に重要な特徴量

図 2 に示す様に、全モデルに関して、DS すなわち停車位置を示す値の重要度が最大となった。停止線までの距離が近い場合は 2 つの状態が予想される。第一に、交差点に停車している車両は少なく、信号変化後すぐに発進できるため、停車時間は短くなる状態、第二に、右折待ちで停車時間が長くなる状態である。一方、停止線までの距離が遠い場合は、交差点に停車している車両が多く、信号が青になってから発進するまでに時間がかかり、停車時間が長くなると考えられる。

また、停車時間分布ごとの 3 モデル共通で、直前の速度または速度変化量の重要度が上位 5 位以内に入った。停車直前の速度が大きい場合には、赤信号に遷移する直前に停車して、停車時間は長くなる可能性が高い。一方で、速度が小さい場合は、すでに赤信号になっていて、停車の準備がされた状態であり、停車時間は短いと予想される。よって、停車位置や直前の速度は停車時間の予測に重要であると考えられる。

Class 1 < Class 2 のモデルに関しては、DT-75 に関連した特徴量の重要度が大きくなった。旅行時間が長くなるのは、信号による停車の可能性が高いため、DT-75、DT-75V、DT-75M は、信号待ち時間の代替値として考えられる。Class 2 の停車の要因として、信号待ちが多いと予想されるため、DT-75 に関連した特徴量の重要度が大きくなったと考えられる。また、赤信号の表示時間は道路区間によって異なるため、停車時間の予測に必要な特徴量だと推測される。

さらに、Class 1 > Class 2 のモデルに関しては、他のモデルと比べて、DT-50V の重要度が大きくなった。旅行時間の中央値は通勤などによる時間帯によって発生する渋滞に影響を受け、また Class 1 の短い停車は渋滞による停車が含まれると予想されることから、この結果は妥当だと考えられる。以上より、停車時間の予測には、停車位置、直前の速度に加えて、赤信号の表示時間や渋滞状況などリアルタイム情報が重要だと考えられる。

ところで、平日と休日の交通量の違いが停車時間の予測に影響を与えると考え、DH を特徴量に加えた。交通量に関して集計した結果、予想通り休日の交通量は平日の交通量の区間平均 1.41 倍 (区間最大 1.80 倍、区間最小 1.18 倍) であった。一方、図 3 より、停車時間の分布の傾向には平日と休日では大きな差はみられなかった。また、全てのモデルに関して、DH の重要度は、直前の速度や旅行時間に比べて小さい。以上より、平日に比べて、休日の方が交通量が多いが、停車時間の予測には、上述の特徴量の方が重要であると考えられる。

5 おわりに

本研究は、特殊なハードウェアが不要な状態で、様々な要因の停車時間の長さを予測することを目標とする。道路区間ごとに停車時間予測モデルを構築するのは多大なコストがかかる。そこで本稿では、初期検討として、道路区間共通モデルを構築するための特徴量を探索した。まず、車両位置情報と道路の地理情報を用いて停車時間の長さを 3 クラスに分類する問題とし

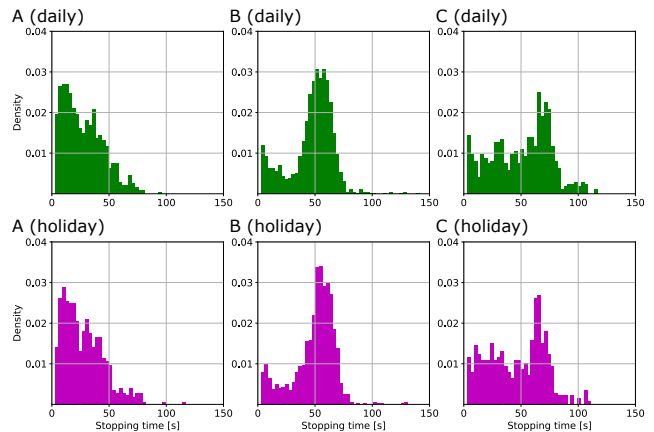


図 3 平日と休日における道路区間ごとの停車時間比較 (道路 3 区間抜粋)。横軸は停車時間、縦軸は累積相対度数を示す。

て設定した。次に、道路全 26 区間で発生する停車を全て含む全データセット、道路区間ごとのサンプル数を揃えるためアンダーサンプリングしたデータセット、停車時間の分布で 3 分割したデータセット、それぞれに対して計 5 個のモデルを構築した。その結果、停車開始地点から道路区間の端までの距離の調整値や直前の速度の重要度が大きくなった。したがって、停車時間の予測に重要な特徴量は、停車位置、直前の速度などリアルタイム情報だと考えられる。一方、どのモデルも道路区間ごとのサンプル数と正解率に相関がみられ、全道路区間に共通性のあるモデルは表現できていない。より精度の高い汎用モデルを構築するために、モデル共通で重要度が大きい停車位置や直前の速度遷移を詳細に分析し特徴量に組み込むことが必要だと考えられる。

今回使用した FCD に関して

今回使用した FCD は、Honda 車に搭載の純正カーナビゲーションシステムから携帯電話網を通して収集されたデータを Honda Total Care 会員規約およびインターネット・プレミアムクラブサービス利用規約等に基づき、個人を特定できないよう加工および統計処理して分析に利用している。本稿では、研究の対象となる道路区間上を走っている FCD を抽出して分析しており、個々の行動履歴を推測できるものではない。

参考文献

- [1] Regan, Michael A., et al, "Driver distraction: Theory, effects, and mitigation", CRC Press, (2008).
- [2] 画像表示装置の取り扱いについて 改訂第 3.0 版, http://www.jama.or.jp/safe/guideline/pdf/jama_guidelines_v30_jp.pdf (最終閲覧: 2019 年 5 月 22 日) .
- [3] 一般財団法人道路交通情報通信システムセンター: 信号情報活用運転支援システム (TSPS), <http://www.vics.or.jp/know/service/tsps.html> (最終閲覧: 2019 年 6 月 19 日) .
- [4] 阿部幸康ら, "北海道における信号交差点のある 2 車線道路の旅行速度", 交通工学, Vol.21, (1986).
- [5] 小林正明ら, "交通ネットワークにおける OD 旅行時間の解析と算定" 情報処理学会研究報告高度交通システム (ITS), (2002).
- [6] 清水光ら, "M-021 都市道路網の動的交通情報を用いたリンク旅行時間の一推定", 情報科学技術フォーラム講演論文集, (2011).
- [7] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python" Journal of Machine Learning Research 12, (2011).