

数理科学実践研究レター 2019–15 November 28, 2019

カーレースのデータにおける変数重要度の評価

by

野島 遼



**UNIVERSITY OF TOKYO**  
GRADUATE SCHOOL OF MATHEMATICAL SCIENCES  
KOMABA, TOKYO, JAPAN

# カーレースのデータにおける変数重要度の評価

野島 遼<sup>1</sup> (東京大学大学院数理科学研究科)

Ryo Nojima (Graduate School of Mathematical Sciences, The University of Tokyo)

## 概要

レーシングカーの加速度に対して、どのような因子が重要かを走行データから考察した。具体的には、走行データに対して加速度を予測するランダムフォレスト回帰を行い、変数重要度を評価した。

## 1 はじめに

カーレースにおいて、レースタイムを短縮することは一番の目標である。そのために、車体に関する様々な因子がどのように車体の運動を決定しているかを理解することは重要である。一つのアプローチとして、車体の運動を物理的にモデル化することがあげられる。例えば、車体の加速度について考えてみると、エンジンの駆動力、空気抵抗、摩擦力などの車体への外力となる因子を用いた運動方程式によりモデル化できるはずである。実際の走行を高い精度で再現できるモデルを作ることができれば、そのモデルを解析することで「タイムを短縮するためにはどう改善すべきか」といった問題に取り組むことができるであろう。

しかし、車体に関する現象のモデル化には試行錯誤や高度な事前知識が必要となることも多い。加速度の例で考えてみると、何を外力項として含めるべきか、その関数形はどのようなものであるべきかなどを考える必要がある。

ところで、車体にセンサーなどを取り付けることによって、刻一刻と変化する車体に関する様々なデータを取得することができる。それらを用いることで、モデル化することなしに重要な因子を発見したり、あるいはモデル化のための事前知識を得ることはできないだろうか。本研究では、実際の力学を想定しやすい車体正面方向への加速度を取り上げ、ランダムフォレスト回帰を用い、変数重要度を見ることが上で述べたようなアプローチを試みた。

## 2 検討内容

本研究では、車体の正面方向への加速度を決める重要な因子は何なのかという問題に取り組んだ。運動方程式より、ある時刻  $t$  での正面方向への加速度を  $a_t$  とすると、

$$ma_t = F(x_t)$$

という式が成り立つ。ここで、 $m$  は車体の質量、 $x_t$  は車体に関する量 (アクセルの踏み具合、エンジン回転数、速度など)、 $F(x_t)$  は  $x_t$  の関数である。カーレースでの車体の進行のような複雑な系では、 $x_t$  として考えるべき変数、 $F(x_t)$  の具体的な関数形を決定するのは難しい問題であると思われる。本研究では、 $x_t$  としてどのような変数が重要と考えられるかを実際の走行データから検討した。

## 3 方法

変数  $x_t$  や  $F(x_t)$  が未知の状態なので、具体的な関数形を仮定しないような回帰手法で加速度の予測モデルを作成したい。このような手法では、しばしば予測のための計算がブラックボックスになってしまうという問題がある。今回は回帰手法としてランダムフォレスト回帰を用いた。ランダムフォレスト回帰モデルでは、予測に各変数がどのくらい寄与しているかを計る変数重要度 [1] と呼ばれる量がよく知られており、その値を計算した。

以下に、用いたデータセット、およびランダムフォレスト回帰モデルの作成に用いたデータについて述べる。

<sup>1</sup>nojima@ms.u-tokyo.ac.jp

### 3.1 データセット

今回使用したデータセットは、ある車体がコースをテスト走行したときの様々な数値データを含んでいるものである。コースを細かな区間に分割し、各区間で測定された車体に関する量がデータセットには記録されている。この分析では、1周あたり 817 個の区間に分割したデータを用いた。

### 3.2 用いたデータ

ランダムフォレスト回帰による予測モデルの作成には学習用として4周分のデータを用い、評価用として1周分のデータを用いた。モデル作成には、データセット中の全ての変数ではなく表1の変数のみを予測モデルの入力として使用した。変数の選択にあたっては、相関係数の値や物理的な背景知識などから変数の絞り込みを行った。さらに絞り込んだ変数の中で、同一のデータと思われる組がある場合は片方だけを用いることにした。

表 1: 回帰に用いた変数

変数名	説明
Engine Speed (rps)	エンジン回転速度
Brake Press Fr (Pa)	ブレーキ圧 (フロント)
Speed (m/s)	速度
Load FL (kgf)	前左への負荷
Load FR (kgf)	前右への負荷
Load RL (kgf)	後左への負荷
Load RR (kgf)	後右への負荷
Steeng Angle	ステアリング角度
Throttle Pedal	スロットルペダル踏み角度

### 3.3 予測モデル

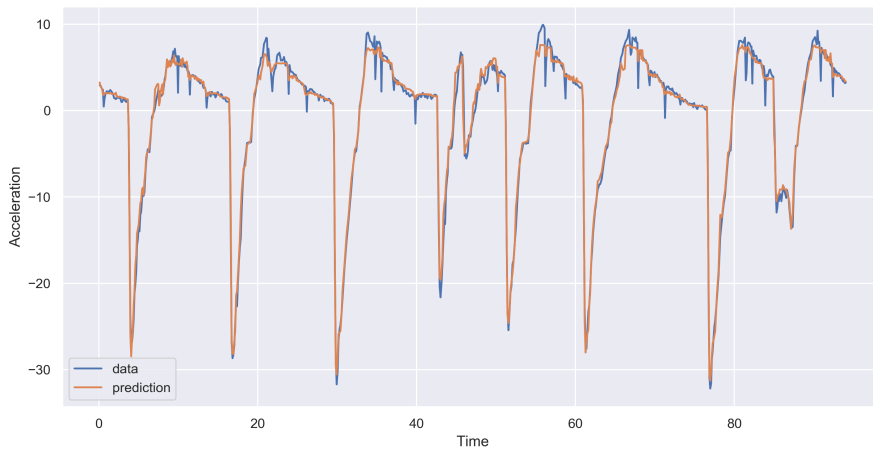
ランダムフォレスト回帰モデルの作成には、Python の scikit-learn ライブラリの RandomForestRegressor を用いた。まず、学習用の4周分のデータで評価指標を平均二乗誤差としてクロスバリデーションを行い、以下の表2のハイパーパラメータについてのグリッドサーチでハイパーパラメータを決定した。その他のパラメータについてはデフォルトの値を用いている。

グリッドサーチの結果、n\_estimators=40, max\_depth=7, max\_features=0.8 の時が最も良かったので、このハイパーパラメータのもとで学習用の4周分のデータ全てを用いてランダムフォレスト回帰モデルを作成した。このランダムフォレスト回帰モデルを利用して、学習に用いなかった1周分の予測を行うと図1のようになり、平均二乗誤差は0.834であった。表1にあげた変数で良く加速度を予測できていると言える。

表 2: グリッドサーチを行ったハイパーパラメータ

パラメータ名	値
n_estimators	40, 60, 80
max_depth	4, 5, 6, 7
max_features	0.6, 0.8, 0.9

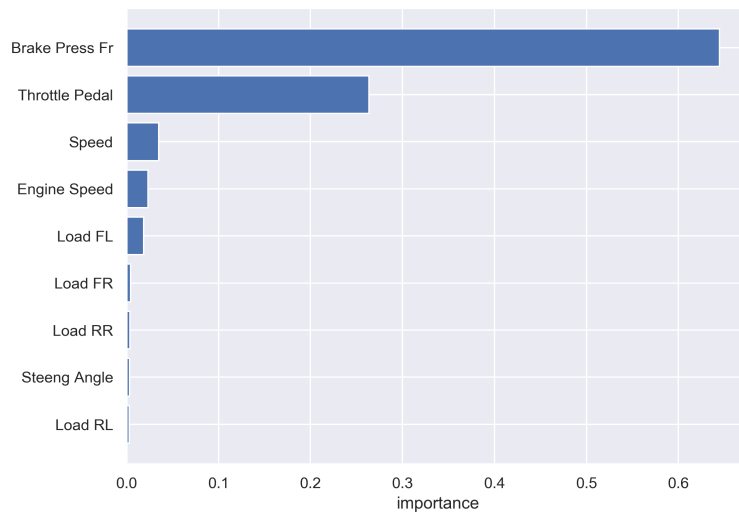
図 1: ランダムフォレスト回帰による予測



## 4 変数重要度

前項で説明したランダムフォレスト回帰モデルでの変数重要度は図 2 のようになった。ブレーキ圧とスロットルペダルの踏み角度が他の変数と比べて重要度が高く、物理的な直感にも整合している。一方、車体への負荷に関する量やステアリング角度については変数重要度は低く、加速度を考える際には無視できる可能性がある。

図 2: ランダムフォレスト回帰による変数重要度



## 5 終わりに

本研究では、ランダムフォレスト回帰により加速度の予測に対する変数の重要性を評価した。加速度については、表 1 で上げた変数により高い精度で予測でき、予測に重要と評価された変数は物理的な直感と整合している。ただし、今回用いたデータセットは周回数としては少なく、限られた状況だけのものであるため、予測モデルが特定の状況に過適合している可能性がある。より多くの走行ケースを含むようなデータセットで再検証する必要があると考えている。

今後の課題としては、より多くのデータがある場合について、このような手法をレースコースの特定の部分に対して試してみることがあげられる。今回はレースコース全てのデータを用いたが、特定の部分での変数重要度を見ることで、走行する際の有用な知見が得られるのではないかと考えている。

また、より一般的な問題にはなるが、ランダムフォレストの変数重要度がデータサンプルやハイパーパラメータに依存してどのような統計的振る舞いをするのかを明らかにすることは数学的にも実用的にも興味深い問題だと考えた。

様々な面で未熟な点のある試みではあったが、走行データを用いることで、走行タイムの短縮に重要な因子を発見したり、車体のパーツを良く調整する方法を確立できるのではないかと考えている。

## 6 謝辞

本課題を提供していただいたアビームコンサルティング株式会社の皆様に感謝いたします。また、多くの有益な助言をくださった東京大学大学院数理科学研究科の田中雄一郎氏に感謝いたします。

## 参考文献

- [1] Breiman, L., *Classification and regression trees*, Routledge, 2017.