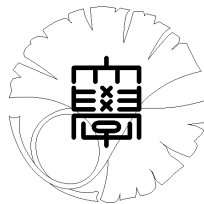


数理科学実践研究レター 2019-1 November 28, 2019

カーレースにおけるコーナーとストレートの差異の解析

by

木下 慶紀



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

カーレースにおけるコーナーとストレートの差異の解析

木下慶紀¹ (東京大学大学院数理科学研究科)

Yoshiki Kinoshita (Graduate School of Mathematical Sciences, The University of Tokyo)

概要

カーレースのデータを解析し、コーナーと直線で速度を決定する要因が変化することを確認した。線形回帰とランダムフォレストを用いて、それぞれの場所で速度に与えている影響が大きい変数を分析した。

1 はじめに

カーレースにおいて車体の挙動は、コーナーと直線で大きく変化する。力学モデルを考える場合、直線部分では進行方向の一次元を考えれば良く、タイヤも進行方向に転がっているため、比較的簡単に運動方程式を与えることができるのに対し、コーナー部分では進行方向と垂直な向きの加速度も考慮する必要があるため、運動方程式は二次元になり、タイヤも滑るので精密な解析は容易ではないと考えられる。またこれらの差異にともなって、早く走るための戦略も直線とコーナーで変わってくると予想される。本研究ではレースカーにセンサーを取り付けてコースを走らせて計測した実データを解析し、コーナーと直線でそれぞれ速度を決定するモデルを構築し、それらの比較を行う。

2 説明変数の選定

本研究で提供されたデータはレーシングカーに同じコースを周回させて計測したデータで、コースをいくつかのセグメントに区切り、そのセグメントにおける各項目（速度、加速度等）の平均値を記録したものである。周回数としては15周分のデータがあるが、この中には流し走行をしているデータも含まれるので、タイムが速かった5周分のデータを使うことにする (LAP 4,8,9,12,14)。またこのデータは非常に多くの項目（1000程度）を含んでいるが、この中にはレース中ほとんど変化しないような物や、何を計測しているかよく分からないものも数多く含まれている。そこで最初にあまり意味を持たなさそうな項目を排除すると、94種類の変数が残った。しかし、これらの中にも共線性があり、速度とはあまり関係の無さそうな変数も残っていた。そのためさらに変数を絞り込むことにした。変数選択を行う方法としては、正則化法や情報量基準を用いる方法などが一般的であるが、これらの手法を実行するためには最初にモデルの候補を決めておく必要があり、また真のモデルの存在を仮定する場合が多い。

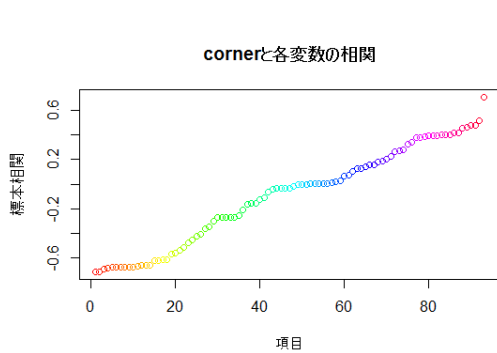


図 1: コーナーと各変数の相関

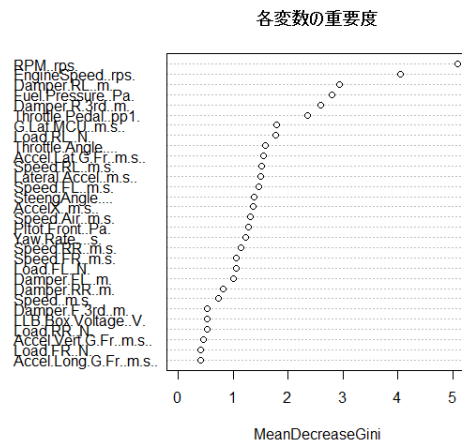


図 2: コーナーを推定するランダムフォレスト

¹yoshikik@ms.u-tokyo.ac.jp

今回の目的は真のモデルを推定することではなく、コーナーと直線でのモデルの違いを比較することであるので、実践的に以下のように変数の選定を行った。まずコーナーで1、直線で0の値を取るダミー変数(以下 corner と呼ぶ)を用意し、その変数と相関の強そうな変数を選択する。図1は corner と各変数の標本相関をプロットしたもので、横軸が変数名、縦軸が標本相関である。図2は corner を推定するランダムフォレストを実行し、ジニ係数に基づいて各変数の重要度 (Importance) を計算してプロットしたもので、縦軸が変数名、横軸が重要度である。本研究では、重要度が高いものや標本相関の絶対値が大きかった変数に、速度に影響の大きそうな変数を加えた以下の10種類の変数を使用することにする。ここで標本相関の絶対値が大きいが、説明変数に選ばなかった物がいくつ

表 1: 説明変数

変数名	標本相関	重要度
横向き加速度	-0.16	0.65
進行方向加速度	-0.017	2.78
垂直加速度	0.26	0.05
ブレーキ圧_前	-0.27	2.04
ブレーキ圧_後	-0.27	1.78
エンジン回転数	-0.72	1.84
ギア比	-0.62	0.05
ステアリング角度	0.27	0.46
アクセルペダル踏み込み	-0.68	4.21

かあるが、これは似たような変数で共線性が見られるものを除外したのと、温度に関する変数を除外したためである。温度に関する変数を除外したのは、速度を決めるモデルを作るときに、速度と温度の因果関係が不明瞭であると判断したためである。今回解析したデータは時系列データであり、独立標本ではないので標本相関が統計的にどの程度有意かを判定するのは容易ではないが、ここで選択した変数はコーナーと直線で変化し、なおかつ車体の速度に影響するという点において、直感的に不自然なものではなかったので、このままこれらの変数を採用し解析を行うことにした。

3 線形回帰モデル

ここでは2節で選んだ説明変数を使い速度を推定する線形回帰モデル (1) を考える。

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon. \quad (1)$$

ここで $X_i (i = 1, \dots, p)$ は表1の説明変数、 Y は目的変数(速度)、 $\beta_i (i = 0, \dots, p)$ はパラメータ、 p は説明変数の種類(ここでは10)、 ε は誤差項であり、 ε は平均0分散 σ^2 の正規分布 $\mathcal{N}(0, \sigma^2)$ に従うと仮定する。通常線形回帰ではパラメータの推定には最小二乗法が使われるが、今回のデータはセグメントごとの平均データなのでその効果を補正することを考える。以下データを $(y_i, \mathbf{x}_i)_{i=1, \dots, n}$ と書くことにする (\mathbf{x}_i は長さ p のベクトル)。また第 i セグメントでは観測が m_i 回行われ、それらの平均が記録されるとする。そして第 i セグメントにおける m_i 個の観測値を $(\tilde{y}_j^i, \tilde{\mathbf{x}}_j^i)_{j=1, \dots, m_i}$ と書く。このとき y_i と \mathbf{x}_i は次のように書ける。

$$y_i = \frac{1}{m_i} \sum_{j=1}^{m_i} \tilde{y}_j^i,$$

$$\mathbf{x}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} \tilde{\mathbf{x}}_j^i.$$

また $(\tilde{y}_j^i, \tilde{\mathbf{x}}_j^i)$ は (1) を満たすのでベクトル表記を用いると

$$\tilde{y}_j^i - \beta_0 - (\beta_1, \dots, \beta_p) \tilde{\mathbf{x}}_j^i = \varepsilon_j^i$$

が成り立つ。ただし $(\varepsilon_j^i)_{i=1, \dots, n, j=1, \dots, m_i} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$ である。したがって

$$y_i - \beta_0 - (\beta_1, \dots, \beta_p) \mathbf{x}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} \varepsilon_j^i$$

となる。ここで $\frac{1}{m_i} \sum_{j=1}^{m_i} \varepsilon_j^i \sim \mathcal{N}(0, \frac{\sigma^2}{m_i})$ であるので誤差分散は m_i に依存することが分かる。そこで m_i によって重みを付けた補正データ $(y_i^\dagger, \mathbf{x}_i^\dagger)$ を $(y_i^\dagger, \mathbf{x}_i^\dagger) = \sqrt{m_i}(y_i, \mathbf{x}_i)$ によって定義する。すると $y_i^\dagger - \beta_0 - (\beta_1, \dots, \beta_p) \mathbf{x}_i^\dagger \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$ となり、通常の線形回帰の文脈に帰着できる。しかし今回使ったデータは m_i が分からない。そこで m_i はそのセグメントを通過するのにかかった時間 t_i に比例すると仮定し、 $(y_i^\dagger, \mathbf{x}_i^\dagger) = \sqrt{K t_i}(y_i, \mathbf{x}_i)$ によって補正データを作成した。ここで K は比例定数。この仮定は観測が時間に対して等間隔に行われていて、各 m_i が小さすぎない状況のもとでは、妥当な仮定であると考えられる。比例定数 K は単位時間あたりの観測回数、すなわち観測間隔の逆数とするのが妥当と考えられる。今回は $K=100$ とした。ただし実際のところ最小二乗法で β_1, \dots, β_p の値を推定する文脈においては次に示すとおり K の値は関係ないことに注意しておく。最小二乗法は次の残差平方和 (RSS) を最小化することで β を推定する。

$$\begin{aligned} RSS &= \sum_{i=1}^n (y_i^\dagger - \beta_0 - (\beta_1, \dots, \beta_p) \mathbf{x}_i^\dagger)^2 \\ &= K \sum_{i=1}^n t_i (y_i - \frac{\beta_0}{\sqrt{K t_i}} - (\beta_1, \dots, \beta_p) \mathbf{x}_i)^2 \\ &= K \sum_{i=1}^n t_i (y_i - \frac{\beta_0^\dagger}{\sqrt{t_i}} - (\beta_1, \dots, \beta_p) \mathbf{x}_i)^2. \end{aligned}$$

ここで $\beta_0^\dagger = \frac{\beta_0}{\sqrt{K}}$ とおいた。最小化において正の定数倍は影響を与えないので結局 K が関わってくるのは β_0 のみであることが分かる。したがってそれぞれの目的変数の係数を比較する分には K の値を意識する必要はない。以上の準備のもとで β の推定をコーナーと直線それぞれで行った結果を

表 2: 線形回帰モデル

説明変数	コーナー	直線	全体
定数項 (β_0)	-233	-376	-404
横向き加速度	-1.5×10^{-2}	-4.8×10^{-2}	-9.2×10^{-3}
進行方向加速度	-5.9×10^{-1}	-6.6×10^{-1}	-6.8×10^{-1}
垂直加速度	2.1×10^{-1}	-2.3×10^{-2}	-5.7×10^{-2}
ブレーキ圧_前	3.3×10^{-5}	3.1×10^{-5}	2.8×10^{-5}
ブレーキ圧_後	-4.4×10^{-5}	-4.4×10^{-5}	-3.9×10^{-5}
エンジン回転数	1.9×10^{-1}	2.7×10^{-1}	2.2×10^{-1}
ギア比	6.9	7.7	7.6
ステアリング角度	-1.6×10^{-2}	-4.2×10^{-2}	-1.8×10^{-2}
アクセルペダル踏み込み	7.8×10^{-2}	5.6×10^{-3}	7.1×10^{-2}

紹介する。表 2 は最小二乗法を用いて実際に β を推定した結果で、コーナーにおける観測データのみを用いて推定した β の推定値、直線における観測データのみを用いて推定した β の推定値及び全ての観測データを用いて推定した β の推定値を表記している。データは LAP 4,8,9,14 のものを用いた。最速 LAP(LAP 12) のデータは 5 節でモデルの評価を行うために使う。この結果から直線とコーナーでブレーキ圧_前やブレーキ圧_後の係数はあまり変化していないのに対して垂直加速度やアクセ

ルペダル踏み込みの係数は大きく変化していることが読み取れる。したがってコーナーと直線でモデルが異なっていると想像できる。

4 ランダムフォレスト

本節では2節で選んだ説明変数から速度を推定するランダムフォレストを実行する。

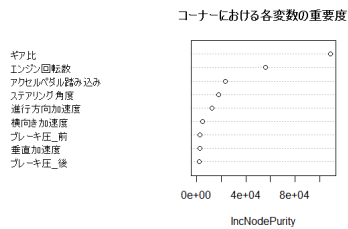


図 3: コーナー

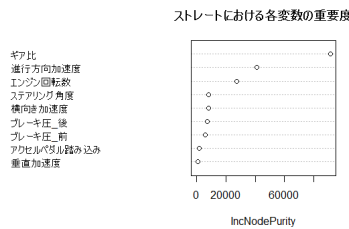


図 4: 直線

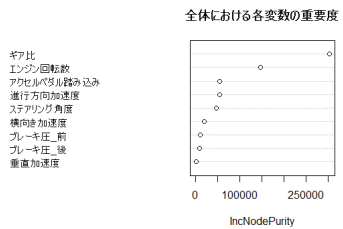


図 5: 全体

図3はコーナーにおける観測データを用いてランダムフォレストを学習したときの各変数の重要度、図4は直線における観測データを用いたときの重要度、図5は全てのデータを用いたときの重要度を表したものである。学習に使ったデータはいずれもLAP 4,8,9,14のものである。直線ではアクセルペダルの重要度が小さいのに対しコーナーでは大きくなっている。また進行方向加速度はコーナーのモデルよりも直線のモデルの方が重要度が大きくなっている。重要度だけでモデルの構造を判断することは出来ないが、コーナーと直線で何らかの違いが生じていることは見うけることが出来る。

5 モデルの比較

本設では3節と4節で構築した線形モデルとランダムフォレストモデルを予測の観点から比較する。具体的にはそれぞれのモデルでLAP 12の速度を予測し、実際の速度との誤差を計算し比較を行う。

表 3: モデル

モデル	コーナー	直線	全体
線形モデル (コーナー)	67	110	91
線形モデル (直線)	187	17	94
線形モデル (全体)	82	31	54
ランダムフォレスト (コーナー)	0.94	18	10
ランダムフォレスト (直線)	37	0.9	17
ランダムフォレスト (全体)	0.75	0.63	0.69
ランダムフォレスト (全体, 半データ)	1.26	1.07	1.16

表3はそれぞれのモデルでコーナー及び直線の速度を予測し予測誤差の二乗平均を表記したものである。最後の行はコーナーと直線両方のデータを使うが使うデータの量を半分にしたものである。線形モデルではコーナーを最も良く記述するのはコーナーのモデル、直線を最も良く記述するのは直線のモデルとなっている。ランダムフォレストではいづれの場面でも全体のモデルが最も良いように見えるがこれは使ったデータの量が多かったことによる可能性が考えられる。そこでコーナーと直線のデータを半分づつ使ったモデルの結果も確認してみる。すると同程度のデータ量ではコーナーを記述するにはコーナーのモデル、直線を記述するには直線のモデルを使うべきであると結論できる。またコーナーのモデルで直線を記述すること及びその逆は精度が著しく悪くなっている。これ

は、コーナーでの速度と直線での速度が異なるメカニズムで決まっていることを示唆している。また全体のデータを使ったランダムフォレストはそれなりに良い精度を出しているが、これは今回使った変数が速度を記述するのにある程度十分な情報を持っていることを示しており、そこがコーナーか直線かが分からなくても速度の推測は十分に可能であると言える。ただしコーナーの情報が不要であることとコーナーと直線で速度のメカニズムが異なるという現象は両立するものであり、その意味でコーナーと直線を分けて考えるという手法自体は一定の効果があると考えられる。

6 終わりに

本研究では車体がコーナーと直線で挙動が変化していることを明らかにし、速度が決まるメカニズムが変わっていることを示唆した。しかし実際にどのようなメカニズムで速度が決まるかを解明するには至らなかった。今後は具体的な速度決定のメカニズムを解明し、より早く走るための戦略を提案することが期待される。またレースにおける究極の目標はラップタイムを縮めることであるが、今回のモデルはセグメント間の相関を無視した解析にとどまっているため、レース全体を見た大域的な最適化という意味では不十分であるといえる。もちろんそれぞれのセグメントを速く走ることが出来れば合計のタイムも向上するが、実際のレースでは、あるセグメントの戦略が次のセグメントのタイムにも影響を及ぼすと考えられるので、局所的なタイムの最適化が必ずしもラップタイムの最適化につながるとは限らない。今後はレースの時系列的な性質も考慮することで、合計のラップタイムを記述するモデルを構築することも課題であると考えている。

謝辞

本研究に際して課題及びデータを提供して頂いたアビームコンサルティング株式会社の皆様に感謝する。またプログラムコーディネーターとしてサポートして頂くとともに多くの助言を頂いた東京大学大学院数理科学研究科特任助教田中雄一郎氏に感謝する。