

# 競馬予想と集団知

07108 入江 洋介 非線形物理学研究室

競馬ファンによる競馬予想と確率モデルによる競馬予想を比較すると、未だに機械の予想は人に負けている。そこで、確率モデルにさまざまな予想を組み合わせることで競馬ファンによる予想に勝つことを目指す。

## 1 競馬予想の作り方

競馬予想に用いるデータとして 1. 出走馬の最終得票率、2. ある競馬情報ベンダーの提供する予想着順 (以下、V モデル)、3. ある競馬新聞の提供する指数 (以下、指数 C)、4. 過去のレースデータをもとに最適化されたロジットモデルのスコア。以上の四つのデータを用いる。競馬ファンは馬の勝率をオッズボードで情報を共有して予想する。オッズは得票率の逆数であり、馬の勝率はその馬の得票率であるという合意が競馬ファンの中で共有される。つまり得票率は競馬ファンの集団知である。2 番目のデータは、出走馬の過去の走破タイムと条件をもとに予想するレースでの走破タイムを予想するニューラルネットワークのモデルである。3 番目のデータは少数の競馬の専門家の意見を平均化したものであるといわれている。専門家の集団知である。

4 番目のデータのロジットモデルについて以下に説明する。H 頭の馬が出走するとして、出走馬の馬番号は  $h \in \{1, 2, \dots, H\}$  で表す。そのうちの h 番目の馬の勝率  $p_h$  を考える。馬は、K 種類の観測可能な属性 (クラス、スピード、成績など) を持ち、 $\vec{x}_h = (x_{h,1}, x_{h,2}, \dots, x_{h,K})$  で表すとす。その馬に騎乗する騎手もまた M 種類の観測可能な属性を持ち、 $\vec{y}_h = (y_{h,1}, y_{h,2}, \dots, y_{h,M})$  で表す。

$$p_h = (\vec{x}_h, \vec{y}_h | \{\vec{x}_h, \vec{y}_h\}_{h=1, \dots, H}),$$

Bolton and Chapman は、この関数形に次のものを用いた。

$$p_h = \frac{1}{Z} e^{V_h} \quad (1)$$

$$Z = \sum_{h=1}^H e^{V_h} \quad (2)$$

$$V_h = \sum_{i=1}^K \theta_i^{V_h} x_i + \sum_{j=1}^M \theta_j^{V_h} y_j \quad (3)$$

$p_h$  の式で分母に表れている Z は、出走する馬での確率の和が 1 になるための規格化因子である。また指数関数の方に乗っている関数  $V_h$  は、馬の価値を表す。そして、レース数 R のデータに対し、各レース  $r \in \{1, 2, \dots, R\}$  での勝ち馬の馬番号を  $h_r^*$  として、次の尤度関数を定義した。

$$\exp(L) = \prod_{r=1}^R p_{h_r^*}$$

関数 L は尤度関数であり最尤法により馬の価値を表す関数  $V_h$  の中の係数  $\theta_i^H, \theta_j^J$  を決定する。ここでは、尤度ではなく AR を最大にするように係数を決定する。

競馬予想のロジットモデルとは、馬の勝率をこれらの属性の関数として表したものである。今回用いた馬の属性は全部で 22 種類、騎手の属性は 4 種類の計 26 種類である。この 26 種類の属性を用いて予想精度が最大になるようにパラメータを決定した。各モデルの Accuracy Ratio (以下、AR) は下表にまとめられている。

予想方法	AR
競馬ファンの予想	67.8 %
V モデル	48.8 %
ロジットモデル	54.5 %
指数 C	65.7 %

表 1 予想方法とその AR

## 2 モデルの包含関係

これらの予想モデルの包含関係を考える。予想モデルを比較したとき、あるモデルがほかのモデルとは別の視点で付加的な情報を加えられるのであれば、モデル同士は異なる価値のあるモデルと考えられる。

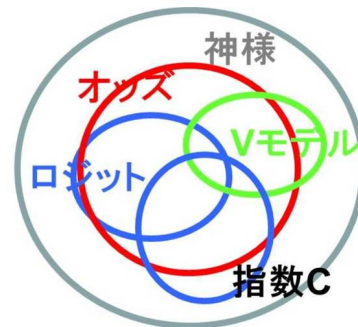


図 1: 各モデルの包含関係。最大の円は神様の予想 (AR=100%)。

図 1 の円の面積はモデルの精度を表しており、面積が大きいほど精度が高い。またこの図では、円が重なった部分は包含関係であることを表している。円の重ならない部分は、相補的な関係にあるということを示している。円の重ならない部分を持つモデルを併用することにより、全体の面積が大きくなりパフォーマンスを上げることが可能である。

各モデルの関係が図 1 の様になっているのならば、それらを組合わせたモデルを作ることによって競馬予想の精度を上げることが期待でき、競馬ファンの予想に勝つことが期待できる。そこでロジットモデルのファクターに 1, 2, 3 の三つのデータを追加し、AR を最大化して最強の予想モデルを作成した。発表ではこうしたモデルの包含関係についても説明する。