

三田祭論文

ファンダメンタル分析とテクニカル
分析を用いた競馬の理論的必勝法

長倉大輔研究会

要 昂輝

植木 樹

久野 主貴

目次

第一章 概要

第二章 ファundamental分析

- 1 節 使用したデータセット
- 2 節 何をどのように分析するか
- 3 節 ランダムフォレスト
- 4 節 特徴量の選定

第三章 テクニカル分析

- 1 節 勝てる馬券とは何か
- 2 節 実証分析—実際のレースでの購入シミュレーション—

第四章 まとめとさらなる課題

参考文献

第一章 概要

これまで競馬は世界中で学術的研究が進められ、統計学に留まらず心理学や経済学といった様々な分野で考察されてきた。特に、確実に払戻金が出費を上回る「必勝法」についての研究は高い関心が集まっており様々な視点から必勝法を導き出す試みが行われている。

馬券の的中率・回収率に関する研究は主に2つに分類される。体重や年齢、過去の成績から着タイムや着順を予測する「ファンダメンタル分析」と、オッズが馬の実力を反映しているという立場から予測する「テクニカル分析」である。多くの先行研究ではこのどちらか一方の手法のみを用いた分析がなされている。しかし、両方の手法を組み合わせ使っている研究は数少ない。本論文では、この2つの分析手法どちらも利用することで回収率100%超、いわば必勝法の導出を試みる。具体的には、「ファンダメンタル分析」についてはランダムフォレストという機械学習の分類手法を用いて、レースに出場する各馬が過去の成績や体格などから3着以内に入るかどうかを予測する。そして「テクニカル分析」では、オッズとファンダメンタル分析で使った機械学習の予測的中率から、払戻金が馬券購入費用を超えるような馬券を選び出す。単純に言うならば、あらかじめ機械学習で複勝圏内の馬を予測しておき、それらの的中率と比べてもオッズが十分に高い場合のみ購入すればよい。

ランダムフォレストを用いたファンダメンタル分析によると、「複勝圏内に入ると予測した馬が実際に複勝圏内に入っている」という的中確率は64.6%だった。この数値から、回収率が1を超えるオッズの最低条件を、ランダムフォレストが何頭を複勝圏内と予測したかによってそれぞれ求めた。最後に、具体的に実際のレースでどれほど活かせるかを2019年に行われた3レースに基づいて検証した。

第二章 ファンドメンタル分析

1 節 使用したデータセット

本論文ではファンドメンタル分析のデータセットとして、JRA-Van Data Lab.よりダウンロードした2016年10月19日から2019年10月19日までのG3以上かつ走破距離が2000メートル以上かつコースが芝またはダートのレース結果を用いる。クラスが未勝利のレース、走破距離が短いレース、障害レースは過去の成績から着順を予測することが困難であるため除いた。取得したデータの総数は1897である。その中から、欠損値と異常値を含むデータを取り除いた1438個のデータでファンドメンタル分析を行う。

2 節 何をどのように分析するか

ファンドメンタル分析は、馬の着タイムの予測に用いられることが多い。予測した着タイムから着順を出し、それに則った馬券（基本的に単勝か3連単の馬券）を購入し回収率が1を超えることを目指す、というものだ。着タイムを予測しているため、ほとんどの研究で取られる分析手法が「回帰」である。結論から述べると、本論文では先行研究も少ない「分類」によるファンドメンタル分析を用いて着順予測をした。理由は2つある。1つ目は、回帰は線形モデル、非線形モデルのどちらにおいても先行研究が多く存在し、その潜在的能力にあまり期待できないからだ。2つ目は、「分類」という手法が後にも出てくるテクニカル分析と非常に相性が良いと思われるからだ。「3着以内に入るか入らないか」という複勝二値分類ならば、その後の馬券購入の組み合わせも複数考えられ、それを工夫する（テクニカル分析）ことでさらなる回収率の増加が期待できる。

「ある特定の馬が3着以内に入るか入らないか」をランダムフォレストという機械学習の分類手法で分析する。

3 節 ランダムフォレスト

ランダムフォレストとは、多様な決定木を多く作り、それぞれの決定木で分類結果を出し、最後にそれらすべてで多数決をとるアルゴリズムである。各決定木を作成する際に、すべての訓練データと特徴量を使うわけではなく一部の訓練データと特徴量のみを使うため、バリエーションが大きいという決定木分類の弱点を克服できる。また、ランダムフォレストは特徴量のスケールに影響されない決定木をベースとしているため、データの前処理の標準化を必要としない、というメリットがある。これは、走破距離や複勝率、斤量といったスケールの大きさが異なる特徴量からなる競馬の複勝分類問題において特に大きなメリットである。さらに、ランダムフォレストでは特徴量の重要度を簡単に可視化することができる。それを基に特徴量を調整し精度を高めていくことも可能だ。

このように、競馬の分類においてはランダムフォレストが最適な手法だと考えられる。次はランダムフォレスト分類するための特徴量を、上記の「重要度の可視化」も利用しながら選定していく。

4 節 特徴量の選定

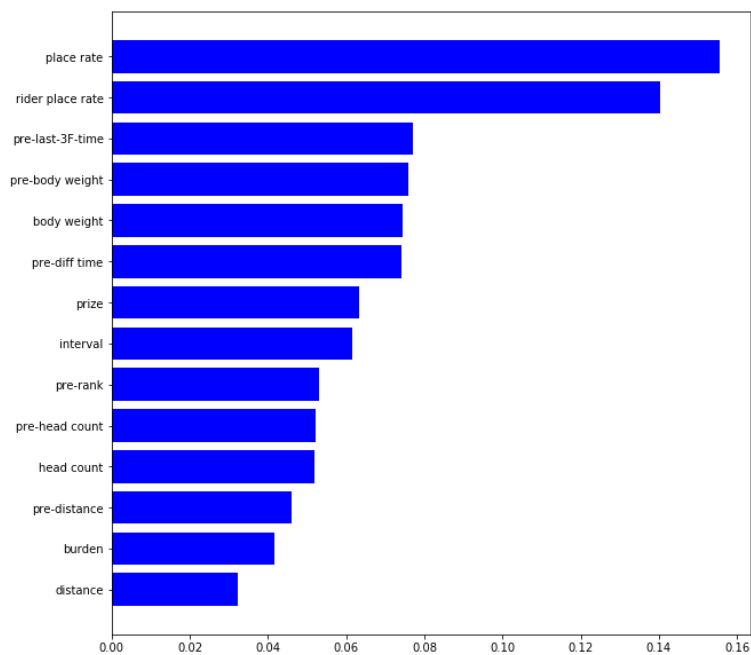
競馬の特徴量の種類は大きく3つに区別することができる。「馬」に関する特徴量、「レース」に関する特徴量、「前走」に関する特徴量である。

まず、「馬」に関する特徴量として、性別、年齢、体重、斤量、確定順位、1着との着差タイム、上がり3Fタイムが取得できた。次に「レース」に関する特徴量として、走破距離、頭数、コースが芝かダートか、馬場状態が取得できた。「前走」に関する特徴量として、上記の特徴量の前走分、前走からの期間的な間隔、前走までの馬の複勝率、前走までの騎手の複勝率が取得できた。これに加えて、レースのクラスを表す指標として、一等賞金額という特徴量もそれぞれ取得した。

これらの特徴量から重要度が低いものを取り除いていき、ランダムフォレストの精度の向上を図った結果、以下の表の特徴量を用いることにした。年齢や性別、馬場状態などは重要度が低く、機械学習の精度を低める要因にもなっていた。また、単勝オッズや人気順という特徴量も含めることは可能だったが、今回は除外した。その理由は2つある。まず、単勝オッズや人気順を加味して重要度の可視化を行うと、これら2つが圧倒的な重要度を示し、他の特徴量の重要度が相対的に相当下がるからだ。さらに的中率も少し下がってしまった。つまり、単勝オッズと人気順の特徴量を含む機械学習はほとんどこの2つの

みで判断してしまい、精度も落ちるのだ。もう一つの理由は、ランダムフォレストでは可能な限り「馬の実力」を反映しなかったからだ。オッズに関してはテクニカル分析で扱うため、ファンダメンタル分析では人気順といった「周りの人間から影響を受ける特徴量」は取り除いた方がよい。簡潔に言えば、ランダムフォレストで「周りの人間たちの考えに左右されない、独立した着順予測を行う装置」を作ることが理想的なのだ。

特徴量名	意味
place rate	馬の複勝率
rider place rate	騎手の複勝率
pre-last-3F-time	前走上がり3Fタイム
pre-body weight	前走体重
body weight	体重
pre-diff time	前走の1着との着差タイム
prize	1等賞金
interval	間隔
pre-rank	前走確定着順
pre-head count	前走頭数
head count	頭数
pre-distance	前走距離
burden	斤量
distance	距離



(特徴量の重要度)

これらの特徴量を用いた結果、テストデータでの正解率は 0.7804878 (約 78%) となった。さらにその正解率を細分化すると、複勝圏内と予測して実際に的中した確率は 64.6% だった。つまり、このランダムフォレスト分類で「3 着以内に入る」という予測が出たとき、それは 64.6% の確率で的中していることを意味する。

$$0.146341463 / (0.080139373 + 0.146341463) = 0.646$$

列1	複勝圏外と予測	複勝圏内と予測
実際に複勝圏外	0.634146341	0.080139373
実際に複勝圏内	0.139372822	0.146341463

第三章 テクニカル分析

1 節 勝てる馬券とは何か

3 章では、2 章で述べたファンダメンタル分析の予測結果とオッズの値を用いて、回収率が 1 を超えるような馬券の買い方を考えていく。まず、回収率とは何であるのかということから考える。回収率とは、的中率を考慮したうえで 1 円分馬券を買ったときに、払戻金がいくらになるのかという値である。今、前述の「ファンダメンタル分析」で A.B.C という 3 頭の馬だけが複勝に入ると予測されたとする。するとランダムフォレストによる予測の分析精度から、馬 A が複勝に入るという予測は 64.6% で的中するといえる。同様に、B が入るという予測も 64.6%、C が入る予測も 64.6% で的中するということが出来るといえる。このことから、A.B.C が全て複勝圏内に入る確率は $0.646 \times 0.646 \times 0.646 = 0.270$ (27%) であるということが出来るといえる。この、A.B.C すべてが 3 位以内に入ることは馬券で言うと 3 連複にあたる。つまり、このことから言えることは A.B.C という組み合わせの 3 連複が予想しているレースで当たる確率は 27% であるということが出来るといえる。そして、現時点で理論を簡単にするために馬券 1 枚を 1 円で買うことが出来ると仮定する。このときに、今予測から導き出された 3 連複のオッズを X とおくと、この馬券が当たることにより、X 円の払戻金を受け取ることが出来ることになる。つまり、X 円が 0.27% で受け取ることが出来る

ということになり、この馬券に期待する払戻金は

$$0.27 \times X = 0.27X \text{ (円)}$$

となる。このお金が馬券を買うことにかかる費用よりも大きければ、馬券を買い続けることにより利益を出すことができることになる。数式でいうと

$$0.27X > 1 \quad \text{または} \quad X > 3.71 \dots$$

を満たすようなオッズがこの3連複の馬券に与えられていればその馬券が買う価値があるということが出来る。

次に、予測の中で複勝に入る馬が4頭出てきた時のことを考慮する。仮にその馬をA.B.C.Dとすると次のようなことが言える。A.B.C.Dがそれぞれ複勝圏内に入る確率は上と同様に0.646である。今、この状況でA.B.Cの3連複を買うとしよう。すると、この馬券が的中する確率は次のように考える。まず、A.B.Cについては上と同様に0.646の確率。しかし、Dについては予想が外れないとこの馬券は当たらないということになる。よって、この3連複の的中率は次のようになる。

$$0.646 \times 0.646 \times 0.646 \times (1 - 0.646) = 0.0954 (9.54\%)$$

よって、上と同様にオッズをXと置くと、

$$0.0954X > 1$$

$$X > 10.5$$

よって、オッズが10.5よりも大きい3連複であれば買う価値のある馬券であると言える。そして、複勝圏内が4頭いるため、この計算を全ての3連複の組み合わせで検証し、当てはまる馬券を買えばいいということになる。そして、このような計算を複勝圏内が5頭、6頭と増えていった時にも行い、当てはまる馬券を買えばいいということになる。

この計算方法の利点としては、実力が均衡している馬が多いのか、それとも離れているのかということの評価できる。単純に3連複に入ればいいだけならば、3連複に入る馬の確率を掛け合わせ、 $0.646 \times 0.646 \times 0.646 = 0.270$ (27%) でよいのである。しかし、あえて複勝圏内にいる馬の予想が外れる確率を計算中に入れることで、レースの特性として実力が拮抗しているレースであるのか、それとも有力馬がいるのかどうかを考慮することが出来るようになる。参考に次の表で予測によって、複勝圏内に入った馬の数と、その時の3連複の的中確率、そして買う価値があると考えられるオッズの最低値をまとめてみる。

複勝圏内の馬の数	3連複の的中率	オッズの最低値
3	0.269586136	3.709389566
4	0.095433492	10.4785016
5	0.033783456	29.600287
6	0.011959344	83.61662995
7	0.004233608	236.2051693

このようなオッズ条件を満たす馬券を購入すれば、回収率が1以上の馬券を購入することが可能になる。また、この結果から言えることはモデルの精度などによって複勝圏内の馬が増えれば増えるほど回収率の高いものを選び出すのが難しくなることがわかる。

複勝圏内の馬の数が2頭だった場合も考慮してみる。その場合、3連複ではなくワイドという購入方法が望ましいと考えられる。ワイドとは、3着以内に入る2頭の組み合わせを的中させる買い方である。上記の3連複の場合と同様に、ファンダメンタル分析で選んだ2頭がどちらも複勝圏内に入る確率は $0.646 \times 0.646 = 0.417316$ である。

$1/0.417316 = 2.396266$ (約2.4倍)

つまり、ワイド馬券のオッズが2.4以上ならば購入し、2.4未満ならば購入しなければよい。

2節 実証分析~実際のレースでの購入シミュレーション~

実際に2019年10月27日の天皇賞において、購入シミュレーションを行ってみた。まず、このレースでのオッズを紹介するのだが、数が多いためモデルに予測された馬に係するもののみ特定する。ちなみに、ランダムフォレストにより選ばれたのは6頭いて、馬番でいうと2、5、6、9、10、15である。

	結果	単勝	複勝
	2019. 10.27 15:5		
1	カデナ	126	7.5
2	アーモンドケイ	1.6	1.1
3	ケイアイノーテック	257.8	16.2
4	スワーヴリチャード	19.1	2
5	アエロリット	20	2.1
6	ユーキャンスマイル	40.6	3.3
7	スティッフエリオ	145.1	11.9
8	マカヒキ	53	6.6
9	ダノンプレミアム	9.5	2
10	サートゥルナーリア	3.4	1.2
11	ゴーフォザサミット	406.1	33.4
12	ドレッドノータス	296.2	18.3
13	ランフォザローゼス	150.9	10.7
14	ワグネリアン	18.7	2.3
15	ウインブライド	148.9	11.3
16	アルアイン	67.9	5.5

3連複			
	2~5~6		56.3
	2~5~9		32.1
	2~5~10		10.3
	2~5~15		400.5
	2~6~9		49.5
	2~6~10		14.1
	2~6~15		612.2
	2~9~10		8
	2~9~15		435.7
	2~10~15		123.9
	5~6~9		429.6
	5~6~10		167.9
	5~6~15		3923.2
	5~9~10		99.5
	5~9~15		2645.6
	5~10~15		1156.3
	6~9~10		112.8
	6~9~15		3828.5
	6~10~15		1262.3
	9~10~15		869.1
表	2019. 10.27 天皇賞 オッズ		

この6頭の馬についての3連複の組み合わせは20通りあり、それぞれの通りは上の表を確認していただければ、それぞれのオッズと共にわかると思う。このオッズと複勝圏内が6頭いるときの回収率1を越える最低オッズの値を比較すると、次の馬券を買うべきであることがわかる。

2-5-15 2-6-15 2-9-15 2-10-15 5-6-9 5-6-10 5-6-15 5-9-10 5-9-15 5-10-15 6-9-10 6-9-15 6-10-15 9-10-15

この14枚が回収率1を超えるという条件に当てはまるものであり、この馬券を全て買えばこの中の1枚が当たった時に利益を出すことが出来るものであるということが出来る。

今回のレースでは当たり馬券である 2-5-9 を選び出すことが出来なかったが、この馬券のオッズは 32.1 であり、予測の精度が上がってもう 1 頭複勝圏内から除くことが出来れば可能性のある馬券だったと評価することが出来る。また、この考え方は伊藤耕介氏の「テクニカル分析の立場をベースとした複勝馬券の的中確率に関する統計モデル」で記載されていた通り、何回も試行し、複数レースの期待値を求めることで必ず回収できるという理論を軸にしているため、予測と考え方、結果全てをとっても評価できるものであるといえる。

では次に、2019 年 4 月 29 日の新潟大賞典 G3 のレースで同様のことをしてみる。今回ランダムフォレストにより選ばれたのは 2 頭（馬番は 6,15）だった。ここでワイドを買うべきかどうか実際のオッズと照らし合わせて検討する。6-15 のワイド馬券のオッズは 7.9 倍、それに対して 3 章の 1 節で述べた買うべき馬券のオッズ最低値は 2.4 だった。このレースにおいてはワイド馬券の購入に踏み切るのが賢明ということになる。実際このレースでは、1 着が 15、3 着が 6 となり見事に的中もしている。

最後に 2019 年 1 月 26 日の愛知杯 G3 についても分析してみる。愛知杯ではランダムフォレストから 3 頭（馬番は 5,8,11）が選ばれた。3 頭の場合は上記の天皇賞の時より非常に分かりやすく、3 連複のオッズが 3.7 以上ならば購入すべきであった。実際の 3 連複のオッズは 90.1 なので迷わず買うべきであるはずだ。しかし、このレースにおいては非常に興味深いことが言える。以下に 5,8,11 からなる 3 連単の組み合わせ（6 通り）とそのオッズの表を載せた。それぞれの場合で、払戻金が 1 円になるために必要な金額も載せた。6 通りすべての金額を合計すると 0.01077 円である。それに対して、3 連複によって払戻金 1 円を獲得するために必要な金額は、 $1/90.1=0.01110$ 円である。つまり、このレースにおいては、3 連複を購入するより、6 通りの 3 連単の組み合わせを購入した方が安上りなのだ。少し本題から逸れたが、どの馬券をどのように購入するかはレースごとによってまだ工夫の余地があると言える。ちなみにこのレースでは 1 着が 5、2 着が 11、3 着が 8 と 3 連複で購入しようが 3 連単で購入しようが見事に的中していた。下の表の通り、5-11-8 の 3 連単は 991.4 倍である。

組み合わせ	オッズ	1円に必要な金額
5,8,11	1206.7	0.000828089
5,11,8	991.4	0.001008675
8,5,11	867.8	0.001152339
8,11,5	504	0.001984127
11,5,8	441.9	0.002262955
11,8,5	283.1	0.003532321

第四章 まとめとさらなる課題

本論文では競馬の必勝法について「ファンダメンタル分析」と「テクニカル分析」の二つの手法を組み合わせ、回収率が1を上回るよう購入する方法を考えた。はじめに、日本中央競馬の2016年10月19日から2019年10月19日の期間における1438個のデータを使用しランダムフォレストでモデルを作成した結果、テストデータにおいて16頭のうち複勝に入ると予測した馬が実際に複勝に入る確率が64.6%となるモデルを構築することができた。このモデルを用いると、実際に複勝に入ると予測された馬の数が3頭ならば複勝馬券の的中率は27.0%、4頭では4通りの三連複馬券それぞれについて9.54%、5頭では10通りそれぞれで3.38%、6頭では1.20%となった。従って、複勝に入ると予測される馬の数が3頭の時は3.71、4頭で10.5、5頭で29.7、6頭では83.7をオッズが上回った際に三連複馬券を購入することによって回収率が1を超え利益が見込めると示された。しかし予測される馬の数が6頭以上の場合には回収率が1を超えるための必要なオッズの最低値が非常に高くなり、実際のレースでの運用は難しくなるため、複勝に入ると予測される馬は5頭以下が好ましいと考えられた。2頭の場合は、ワイドのオッズが2.4以上ならば購入すべきだという結果になった。また、最後の実証のように、レースごとでさらに馬券の購入方法を工夫することができる。全体として、機械学習の精度をさらに向上させ、各レースについて単勝も含めた個別の工夫を考える。そうすればさらに回収率は高まり、必勝法としての価値も高まっていくと考えられる。

参考文献

芦谷政浩 (2012) 「日本の公営競馬における『競馬必勝法』の具体例」、国民経済雑誌、205(6):81-91

伊藤耕介 (2010) 「テクニカル分析の立場をベースとした複勝馬券の的中確率に関する統計モデル、大阪商業大学アミューズメント産業研究所紀要、第 12 号

NUKUI Shun 「AlphaKeiba はなぜ賞金 30 万を獲得てできたのか?」、第一回ウマナリテイクス

足立悠 (2019) 「機械学習のための『前処理』入門」

加藤公一 (2018) 「機械学習のエッセンス - 実装しながら学ぶ Python, 数学, アルゴリズム」、SB クリエイティブ

塚本邦尊、山田典一、大澤文孝、中山浩太郎、松尾豊 (2019) 「東京大学のデータサイエンティスト講座~Python で手を動かして学ぶデータ分析~」