

1. はじめに
2. まだ
3. まだ
4. 提案手法
5. おわりに

# タイトルはまだです

## 英語版タイトル

清水 豪士

**Department of Information Systems Engineering,  
Graduate School of Engineering  
u155016@st.pu-toyama.ac.jp**

**Teams, 9:30-10:00 Tuesday, February 9, 2021,  
Toyama Prefectural University.**

# 1.1 本研究の背景

## 本研究の背景

1. はじめに
2. まだ
3. まだ
4. 提案手法
5. おわりに

- 為替取引や賭け事は一般的にゼロサムゲームと呼ばれる一方が得点すると他方が失点することで、全体の持ち点の和は必ずゼロになるとといったものである。
- 競馬においてはゼロサムゲームの中でも非ゼロサムゲームと呼ばれる参加者の得点と失点の総和がゼロにならない状態に近い。
- そのため、競馬においてはリスクを抑えて利を目指すことが重要である。
- その点が金融などでよく目にするポートフォリオ問題と共通している。
- しかし、そのためにはどの馬が優秀なのかの予想やどのレースならば勝ちやすいのかをデータをもとに予想する必要がある。

# 1.2 本研究の目的

3/24

## 本研究の目的

1. はじめに
2. まだ
3. まだ
4. 提案手法
5. おわりに

- 1 競馬においての馬の強さの数値化
- 2 3着以内に入る馬（払い戻しに該当する馬）を予測する。
- 3 予測したものをもとにリスクを抑え、利益を高めるための馬券購入
- 4 スクレイピングしてきたデータをもとに競馬におけるポートフォリオ問題を解く。

## 2.1 様々な賭け事と共通項

4/24

### 賭け事

賭け事は一般的にはギャンブルといわれ、ルールを基に金銭を賭け、負けた方が賭けた物を失い、勝った方が賭けた物 + 利益を得ることができる。そのため、リスクを伴う賭けになることが多いが、立ち回り次第では利益を最大化できる。

### 賭け事における共通項

一般的にはお金を賭けることが主流である。  
また、お金を賭ける際に人や過去の成績などを基準にして賭けるという点が様々な賭け事において共通している。  
しかし、出走する人数やレースの数に違いはある。

### 賭け事の例

① 競馬、競輪、競艇

## 2.2 賭け事の仕組みと特徴量

5/24

### オッズ

- はじめに
- まだ
- まだ
- 提案手法
- おわりに

- 的中した際の配当を倍率で表したもの。
- オッズを見ることでどの馬が人気なのかを読み取ることができる。
- しかし、人気が高い馬や過去の成績が良い馬などはオッズの変動が少ないが、払い戻しの金額が小さい。

単勝・複勝						
人気	枠	馬番	印	馬名	単勝オッズ	複勝オッズ
1	1	1		エフフォーリア	1.7	1.1 - 1.3
2	8	16		サトノレイナス	5.1	1.6 - 2.3
3	7	13		グレートマジシャン	11.1	2.3 - 3.9
4	5	10		シャフリヤール	11.7	2.3 - 4.0
5	6	12		ワンドフルタウン	16.2	2.5 - 4.4
6	4	8		ヨーホーレイク	18.4	2.9 - 5.2
7	3	5		ティーブモンスター	23.7	4.4 - 8.0

図 1: 2021 日本ダービーにおけるオッズ

特徴量	
共通	番号, 過去の成績, レース場, 距離, 勝率, レースグレード, 天気など
競馬	馬の強さ, 馬の性別, 年齢, 馬体重, 距離適正, 脚質, 血統など

図 2: 特徴量

## 2.3 従来研究の紹介（2章に基づいて）

6/24

### 競馬に対する戦略の最適性の照明

胴元がオッズを設定し,  $m$  頭が出走するレースで 1 着になる馬を賭け師が予想する競馬を考える。

そのもとで賭け師のレース後の資金の成長率が最大となる賭け方を最適化問題として捉え, その最適解を考える, 最適性の証明を行った<sup>1</sup>.

過去の中山競馬場での勝馬の成績のデータを用いて因子分析と主成分分析を行い, 有馬記念の勝馬の予想を行った<sup>2</sup>.

ファンダメンタル分析で出走馬の選別を行い, テクニカル分析を用いて的中率とオッズがともに高い場合の馬券を購入するといった方法で予測を行った.

---

<sup>1</sup>中川 勇斗, “不公平なオッズの競馬に対する戦略の最適性の証明”, 信州大学 工学部 学士論文 2016.

<sup>2</sup>辺見 広大, “主成分分析と因子分析による競馬の勝因の研究”, 大阪工業大学 2009.

## 3.1 賭け事の数理モデル化

7/24

調べ中

ポートフォリオの数式を用いて、その数式の変数には何が入るのかを検討する。

1. はじめに
2. まだ
3. まだ
4. 提案手法
5. おわりに

## 3.2 Bradley-Terry モデルとゲーム理論

8/24

### Bradley-Terry モデル

- スポーツやゲームの勝敗や、アイテムの好き嫌いの一対比較の対象データを用いて強さを表現するモデル。

### 競馬に BT モデルを適用するなら

- 2着馬に注目すると、1着馬に負け、3着以下の馬には勝ったという形になる。

$$P_{ij} = \frac{\pi_i}{\pi_i + \pi_j}$$

図 3: Bradley-Terry モデル

- 要素  $i$  が要素  $j$  に勝利する確率を  $P_{ij}$  とし、 $\pi_i$  は要素  $i$  の強さを表す。

1. はじめに
2. まだ
3. まだ
4. 提案手法
5. おわりに

## 3.3 分析手法

9/24

### アンサンブル学習

- 複数のモデル（学習器）を融合させて 1 つの学習モデルを生成する学習方法。

1. はじめに
2. まだ
3. まだ
4. 提案手法
5. おわりに

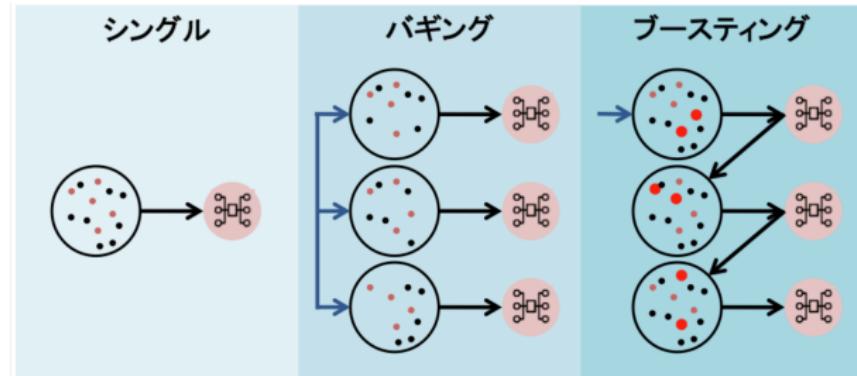


図 4: アンサンブル学習

## 3.3 分析手法

10/24

### ブースティング

1. はじめに
2. まだ
3. まだ
4. 提案手法
5. おわりに

- アンサンブル学習の手法の 1 つ.
- ブースティングは前の弱学習器の結果を次の学習データに反映させるという特徴を持つ.

### 勾配ブースティング

- アンサンブル学習のブースティングを用いた手法.
- 勾配ブースティングはそれぞれの弱学習器の誤差を学習することに最大の特徴を持つ.
- 予測精度は高いが、計算時間が長いという特徴も持つ.

### 3.3 分析手法

#### LightGBM

- 決定木アルゴリズムに基づいた勾配ブースティングの機械学習フレーム。
- Leaf-wise という手法を採用しており、同じ勾配ブースティングを用いた機械学習フレーム XGboost の Level-wise 比べて訓練時間が短くなる傾向がある。

#### 重回帰分析

- 1つの目的変数を複数の説明変数で予測するもの。

## 4.1 競馬予測のプラットフォームの構築

12/24

### netkeiba のスクレイピング

netkeiba.com のスクレイピングを行う.

- netkeiba の数年分のレース結果を取得した.
- 情報として、着順、騎手、人気順、オッズなどの情報

馬体重	調教師	馬主
460(+2)	[東]金成貴史	ニットー商事
432(+4)	[東]林徹	ノルマンディ
458(0)	[東]萱野浩二	ミルファーム
418(0)	[東]星野忍	内田玄祥
536(-2)	[東]加藤征弘	J. ウー

図 5: スクレイピングの一部抜粋

## 4.2 データ分析

1. はじめに
2. まだ
3. まだ
4. 提案手法
5. おわりに

- スクレイピングで取得したデータをもとに馬の「強さ」を算出する。
- その「強さ」を特徴量の1つとして捉え、データ分析を行う

例) 着順を目的関数として、着順以外の特徴量が着順に対してどれくらいの重みを持っているのかを調べる。

### イメージ手順

- 1 特徴量を取得し、それらを用いて重回帰分析を行いどの特徴量が重要なのかを調べ、特徴量の選別を行う。
- 2 選別した特徴量をもとに、LightGBM を用いて3着以内になる確率を予測する。

## 4.3 未定

14/24

ポートフォリオ問題についての解法を書くかも.

1. はじめに
2. まだ
3. まだ
4. 提案手法
5. おわりに

# 5. おわりに

## まとめ



1. はじめに
2. まだ
3. まだ
4. 提案手法
5. おわりに

## 今後の課題

- スライドの完成度を高める
- スクレイピングの際レース結果のみだけでなく、馬の情報部分も取りたい。
- 重回帰分析をかける際にスピード指数を算出する。また、偏回帰係数を見てどの偏回帰係数が必要なのかを吟味する。(イメージとしてタイムを目的変数とする。)

## 進捗

1. はじめに
  2. まだ
  3. まだ
  4. 提案手法
  5. おわりに
- ポートフォリオ問題について調べた
  - ポートフォリオ問題を競馬に当てはめたときの一例の紹介
  - スクレイピングしたときに出力される csv ファイルの整形

## ポートフォリオ問題

1. はじめに
2. まだ
3. まだ
4. 提案手法
5. おわりに

- ポートフォリオとは自分の持っている金融資産の按分率のこと。
- リスク資産（株式など）へ投資を行うとき、リスクを減少させるためにはポートフォリオを組み、複数資産への分散投資をすることが推奨される。
- 一般的にリスクが最小で平均収益率が最大となるポートフォリオを得ることはできない。
- リターンとリスクの間にはトレードオフの関係があり、それを考慮して最適なポートフォリオを選択する必要があり、そのため用いられる数理計画モデルをポートフォリオ最適化モデルという<sup>3</sup>。

---

<sup>3</sup> 枇々木 規雄, “ポートフォリオ最適化入門”,  
オペレーションズ・リサーチ, 2016年6月号, pp. 335-340, 2106.

## 金融市場におけるポートフォリオ問題

- はじめに
- まだ
- まだ
- 提案手法
- おわりに

- 市場に資産が  $n$  個あるとき、資産  $i$  の収益率を  $R_i$  とし、 $i$  にはリスクがあるとする。
- 自身の資金の  $100w_i\%$  を  $i$  に投資するものとし、投資比率を表す  $w_i$  をポートフォリオと呼ぶ。 $(\sum_{i=1}^n w_i = 1$  を満たす)
- $w_i$  による投資の収益率を  $R_p$  とする。
- ポートフォリオの期待収益率： $E(R_p) = \sum_{i=1}^n w_i E(R_i)$
- ポートフォリオの収益率の分散： $\sigma_p^2 = \sum_{i=1}^n \sum_{j=1}^n w_i w_j Cov(R_i, R_j)$
- この分散を最小化することが目的になる

## 2つの資産からなるポートフォリオ

期待収益率： $E(R_p) = w_A E(R_A) + w_B E(R_B)$

収益率の分散： $\sigma_p^2 = w_A^2 \sigma_A^2 + w_B^2 \sigma_B^2 + 2w_A w_B \rho_{AB} \sigma_A \sigma_B$

ただし、 $\sigma_i = \sqrt{Var(R_i)}$ ,  $\rho_{ij} = Corr(R_i, R_j)$

1. はじめに
2. まだ
3. まだ
4. 提案手法
5. おわりに

- ポートフォリオ問題における期待値, 分散を競馬の場合に置き換えると, それぞれ馬券当選時のリターンと馬のオッズにあたる.
- 馬券1と馬券2のあたる確率をそれぞれ

$$P(X_1 = 1) = p_1, P(X_2 = 1) = p_2 \quad (1)$$

と表す.  $X$  は確率変数で  $X = 1$  で当たった時,  $X = 0$  で外れを表す.

- 馬券に対しての投資割合をそれぞれ  $Z_1, Z_2$  とすると

$$Z_1 + Z_2 = 1 \quad (2)$$

が成立する.

- それぞれのオッズを  $r_1, r_2$  とすると, リターンは

$$r_1 Z_1 = y_1, r_2 Z_2 = y_2 \quad (3)$$

と表すことができ,  $(Z_1, Z_2)$  の投資に対するリターンを  $Y$  とすると

$$Y = y_1 X_1 + y_2 X_2 \quad (4)$$

のようにリターンを確率変数で表すことができる.

# 競馬におけるポートフォリオの続き

- リターン  $Y$  に対しての期待値は

$$E[y_1X_1 + y_2X_2] = y_1E[X_1] + y_2E[X_2] = y_1p_1 + y_2p_2 \quad (5)$$

と表すことができ、期待値に対しての分散は

$$\begin{aligned} E[(y_1X_1 + y_2X_2 - (y_1p_1 + y_2p_2))^2] &= E[(y_1(X_1 - p_1) + y_2(X_2 - p_2))^2] \\ &= y_1^2E[(X_1 - p_1)^2] + y_2^2E[(X_2 - p_2)^2] + 2y_1y_2E[(X_1 - p_1)(X_2 - p_2)] \end{aligned} \quad (6)$$

ここで、

$$E[(X_1 - p_1)^2] = \sigma_1^2, E[(X_2 - p_2)^2] = \sigma_2^2, E[(X_1 - p_1)(X_2 - p_2)] = \sigma_{12}$$

とおくと

$$E[(y_1X_1 + y_2X_2 - (y_1p_1 + y_2p_2))^2] = y_1^2\sigma_1^2 + y_2^2\sigma_2^2 + 2y_1y_2\sigma_{12} \quad (7)$$

となる。

- (5) の期待値を定めた時、(2) の制約のもとで (7) を最小化する問題を競馬においてのポートフォリオ問題の最適化問題と捉えることができる。

# スクレイピングの変更点

21/24

## csv の整形

- はじめに
- まだ
- まだ
- 提案手法
- おわりに

- もともとの csv への出力内容では不要な部分があったため、その部分を削除し、使いやすくした。

馬体重	調教タイム	厩舎コメント	備考	調教師
490(+2)				[東]中川公成
494(-2)				[東]加藤征弘
536(+8)				[東]中館英二
462(-4)				[東]佐藤吉勝



馬体重	調教師	馬主
460(+2)	[東]金成貴史	ニットー商事
432(+4)	[東]林徹	ノルマンディ
458(0)	[東]萱野浩二	ミルファーム
418(0)	[東]星野忍	内田玄祥
536(-2)	[東]加藤征弘	J. ウー

- 画像のような不要なデータも取ってきてしまうので、それを削除した。

## 合成オッズ

1. はじめに
2. まだ
3. まだ
4. 提案手法
5. おわりに

- 複数の買い目を合成した期待値（オッズ）のこと。
- 馬券の複数の買い目のオッズを計算することで、1つの馬券のように何倍のリターンが見込めるかがわかる。
- トリガミにならないように資金配分をするのに必要なのが合成オッズという考え方。

## 合成オッズ

- 1 それぞれのオッズの逆数を計算する。
- 2 1ででてきた数字を足す。
- 3 足した数字の逆数を計算する。
- 3で出てきた数字が合成オッズを表す。

# 使えそうな考え方

23/24

## オッズの歪み

1. はじめに
2. まだ
3. まだ
4. 提案手法
5. おわりに

馬番	馬名	単勝	馬単
1	モズベッロ	68.8	72.9
2	サリオス	5.8	5.6
3	アーデントリー	202.0	234.4
4	ブラヴァス	112.9	98.1
5	ペルシアンナイト	111.3	111.8
6	ワグネリアン	49.8	45.4
7	コントレイル	1.8	1.5
8	レイパパレ	12.2	12.2
9	クレッセンドラヴ	95.1	107.3
10	カデナ	182.5	243.6
11	ハッピーグリン	283.0	513.2
12	グランアレグリア	2.8	3.0
13	アドマイヤビルゴ	77.6	75.3

- 単勝を 1 点買うのと馬単で頭固定で相手全通り買うのは的中率が同じである。
- しかし、単勝オッズと馬単合成オッズとが違っている馬がレースには存在する。
- このような不整合のことを「オッズの歪み」という

# まとめ

24/24

## まとめ

- CSV の整形
- ポートフォリオについてのサーベイ

1. はじめに
2. まだ
3. まだ
4. 提案手法
5. おわりに

## 課題

- オッズを用いるポートフォリオにオッズ以外の要素を足す方法
- 合成オッズの考え方をポートフォリオに使うことができるのかの検討
- スクレイピングの際に血統などの情報を取ってくるかの検討