

# *Sumo Predict Diaries*

*Nov. 2019*



## <目次>

1. 分析の前提

2. 予測モデル

3. データでみる相撲

4. 今後の予定



# <目次>

**1. 分析の前提**

**2. 予測モデル**

**3. データでみる相撲**

**4. 今後の予定**



# そもそも、何で相撲予測をやってるの？

相撲予測を始めた経緯/目的:

- データ分析を通じたドメイン知識の獲得【◎】
  - 素人であるほど、先入観無く始められるメリットも
  - この発想はまさに『データドリブン思考』(の一例)
- 機械学習による予測精度の検証【○】
  - 場所開催中は15日間、毎日検証が行える
  - 隔月開催の為、モデルフラッシュアップの時間も確保可能
- 競馬予測の為の事前準備【◎】
  - いきなり、競馬予測に着手するのは難しい為、簡単なテーマで肩慣らしをしておきたかった
  - ルールが分かりやすい。国民的(?)スポーツ



# 大相撲とは力士が一对一の素手で行う 真剣勝負の格闘技

大相撲とは？(1/2):

## < 取組に関するルール >

- 取組は両者の暗黙の了解で決まる「立ち合い」で開始される
- 手で使って良いのは手のひらだけ。  
拳で殴ったり、肘打ちなどは反則となる

- 勝敗は以下のいずれかによって決まる
  - (a) どちらか一方の体が土俵外に着地する
  - (b) どちらか一方の足の裏以外が着地する
  - (c) 反則行為を行った場合(8つの禁じ手)



# 大相撲の醍醐味は、決まり手の多さ(?)と『小よく大を制す』の場面が見られること

## 大相撲とは？(2/2):

- **決まり手: 全部で82手**

{ 基本技(7), 投げ手(13),  
掛け手(18), 反り手(6),  
捻り手(19), 特殊技(19) }

- **非技: 全部で5手**

{ 勇み足, つき手, 踏み出し,  
腰砕け, つきひざ }

### ◆ 決まり手上位10手

No	決まり手	構成比(%)	累積構成比(%)
1	寄り切り	27.6%	28%
2	押し出し	22.3%	50%
3	叩き込み	8.9%	59%
4	突き落とし	6.0%	65%
5	上手投げ	4.4%	69%
6	引き落とし	3.9%	73%
7	突き出し	3.2%	76%
8	寄り倒し	3.0%	79%
9	押し倒し	2.8%	82%
10	送り出し	2.7%	85%
-	その他	15.0%	100%

※データ期間: 2013/01~2019/09

- **大相撲の特徴: 体重による階級分けがない!**

巨漢力士を体重が半分ぐらいの小兵力士が倒す、  
といった『小よく大を制す』場面が見られることも!



# 今場所の日々の取り組みの勝敗を予測

## 分析の前提:

- 予測対象(目的変数):
  - 場所中の日々の取り組みの勝敗を予測する
- 番付ランク: 幕内・十両のみ (上位2階級)
  - { 「幕内」「十両」「幕下」「三段目」「序二段」「序ノ口」 }
- テータ期間: 直近2.5年
  - 2017年1月場所～2019年9月場所 (隔月)
- 先場所までのテータを用いて今場所の勝敗を予測
  - つまり、今場所中の前日までのテータは利用しない
  - 今後は、今場所中のテータの利用も検討予定



# 学習用と予測用で異なるデータソースを選定

## データソースの選定:

### < 学習用 / 予測用データの要件 >

- ▣ **【学習用データ】**  
過去分のデータが十分に蓄積されていること
- ▣ **【予測用データ】**  
リアルタイムにデータ更新がされていること

### < データソース候補 >

- **[A] 相撲協会: ▣ 【予測用データ】**  
<http://www.sumo.or.jp/>
- **[B] 相撲シファレンス: ▣ 【学習用データ】**  
<http://sumodb.sumogames.de/Default.aspx?l=j>
- **[C] 相撲星取表(250年以上にわたる大相撲の記録サイト):**  
<http://sumo-hositori.com/link/link.html>



# <目次>

1. 分析の前提

2. 予測モデル

3. データでみる相撲

4. 今後の予定



# 今回は(a)(b)2つの予測モデルを構築

## 予測モデル概要:

- (a) 機械学習モデル:

- { 勝ち=1, 負け=0 }の2値を機械学習モデルで予測
- LightGBMモデル

- (b) 戦闘力推定モデル (※次スライドも参照):

- 過去の対戦成績から各力士の強さ(Intensity)を推定
- Bradley-Terryモデル (1952年に提案された古い統計モデル)

- (c) (a)と(b)のハイブリッドモデル:

- ハイブリッドにすると精度が上がるか?
- 現在は未対応



# Bradley-Terryモデルは、対戦型のスポーツにおいてチームの強さを推定する為のモデル (参考)Bradley-Terryモデルについて:

パウル君



- 野球, サッカー, 相撲, 囲碁・将棋などの勝敗データの利用  
例.) W杯の勝敗予測(パウル君!)

- BTモデルの発展系のモデルもたくさんある(らしい)  
例.) 引き分けや、時系列を考慮するもの etc

- $p_{ij}$ : プレイヤー  $i$  がプレイヤー  $j$  に勝つ確率.  $V_i = \log(\pi_i)$ :  $i$  の戦闘力
- $V_i - V_j = \log(\pi_{ij}) - \log(\pi_{ji}) = \log\left(\frac{\pi_i}{\pi_i + \pi_j}\right) - \log\left(\frac{\pi_j}{\pi_i + \pi_j}\right) = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right)$
- 戦闘力の差が対数オッズ比になるとする. (差が大きいと確率が大きくなる)
- 全プレイヤーの対に関して独立性を仮定して積をとった尤度関数で最尤法.
- ただし, パラメータの識別のために制約が必要.
- よくある制約は, 戦闘力パラメータの和が0か, 戦闘力パラメータのどれかが0.



# 予測ツールを作成し自動レポート

## 予測ツール概要(1/5):

Hyperlink

[ [https://ryutoro-galois.github.io/sumo-predictor/PredResult\\_sumo\\_targetYM\\_201911.html](https://ryutoro-galois.github.io/sumo-predictor/PredResult_sumo_targetYM_201911.html) ]

- レポート出力時刻(daily):
  - 7時:予測結果 (**01\_pred**)
  - 18時(取り組み終了後): 的中率 (**02\_backtest**)

### <ツール機能>

- 勝敗予測 (a.機械学習モデル)
- " (b.戦闘力推定モデル)
- 力士の強さランキング (b.戦闘力推定モデル)
- モデル別の中状況 (a, b各々)



# 予測ツールを作成し自動しポータリング

## 予測ツール概要(2/5): 各モデルによる勝敗予測 (01\_pred)

取り組み開始前に予測結果を出力し、中継を待つ

### 1-a. 勝敗予測 (機械学習モデル)

9日目 (11/18) 8日目 (11/17) 7日目 (11/16) 6日目 (11/15) 5日目 (11/14) 4日目 (11/13) 3日目 (11/12)  
2日目 (11/11) 1日目 (11/10)

【9日目 (11/18) : 取り組み数=31】

Show 10 entries

取組順 (last)	力士X	力士Y	勝者予測	予測確率 (%)	予測一致	勝者結果	決まり手
			All			All	All
1	琴勇輝	白鵬	【白鵬】	80.5	.	.	.
2	貴景勝	北勝富士	【貴景勝】	59.5	.	.	.
3	御嶽海	玉鶯	【御嶽海】	58.5	.	.	.
4	阿炎	竜電	【阿炎】	63.5	.	.	.
5	隠岐の海	遠藤	【隠岐の海】	58.6	.	.	.
6	大栄翔	朝乃山	【大栄翔】	50.8	.	.	.
7	妙義龍	明生	【明生】	50.5	.	.	.

予測結果

勝敗実績(空欄)



# 予測ツールを作成し自動レポーティング

## 予測ツール概要(3/5): 各モデルによる勝敗予測 (02 backtest)

勝敗実績と予測結果を照合し、的中状況を確認

### 1-a. 勝敗予測 (機械学習モデル)

9日目 (11/18) **8日目 (11/17)** 7日目 (11/16) 6日目 (11/15) 5日目 (11/14) 4日目 (11/13) 3日目 (11/12)  
2日目 (11/11) 1日目 (11/10)

【8日目 (11/17) : 的中率=50.0% (16/32)】

Show 10 entries

予測結果

勝敗実績

取組順 (last)	力士X	力士Y	勝者予測	予測確率 (%)	予測一致	勝者結果	決まり手
			All			All	All
1	玉鶯	白鵬	【白鵬】	76.5	○	【白鵬】	寄り切り
2	宝富士	高安	【高安】	79.3	×	【宝富士】	不戦
3	貴景勝	明生	【貴景勝】	61.8	○	【貴景勝】	押し出し
4	御嶽海	碧山	【御嶽海】	62.0	○	【御嶽海】	寄り切り
5	阿炎	朝乃山	【阿炎】	53.0	×	【朝乃山】	押し出し
6	北勝富士	遠藤	【北勝富士】	61.6	×	【遠藤】	押し出し
7	大栄翔	隠岐の海	【大栄翔】	56.0	○	【大栄翔】	押し出し

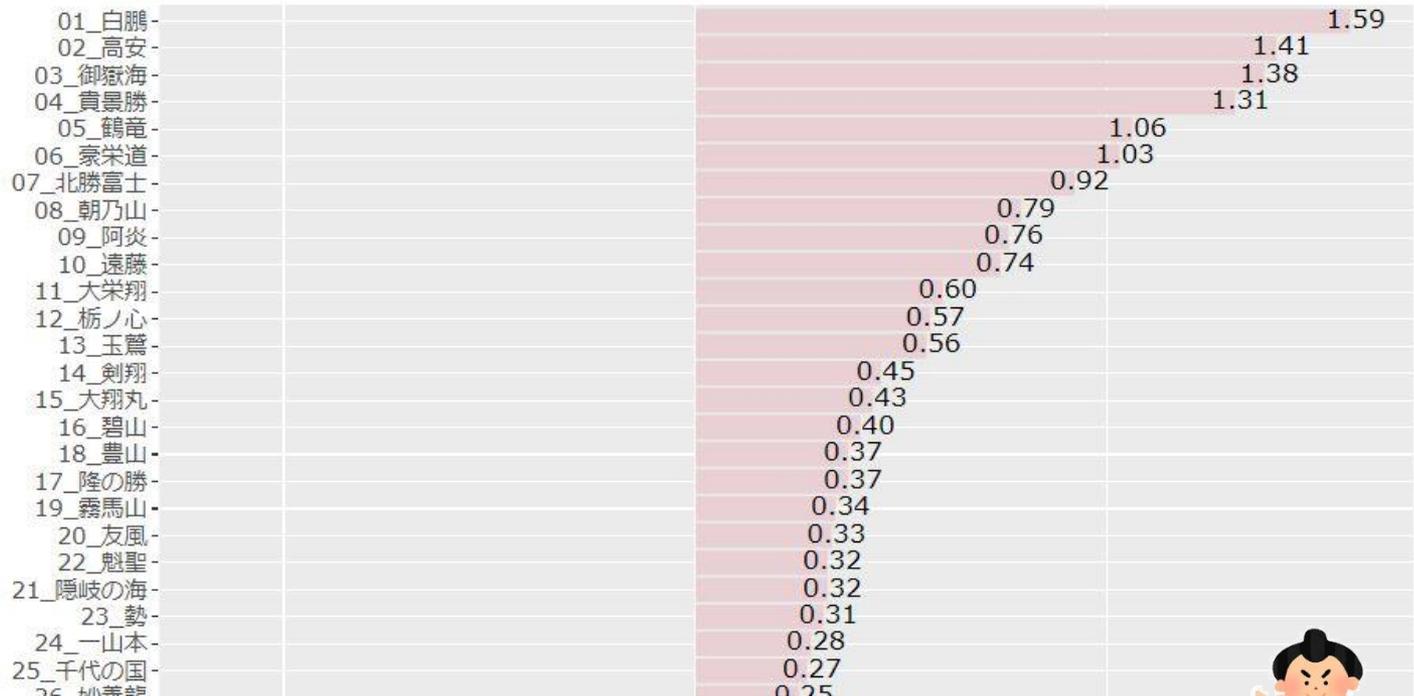


# 戦闘力推定モデルを用いて力士の強さを可視化 予測ツール概要(4/5): 力士の強さランキング (戦闘力推定モデル)

## 4. 力士の強さランキング

令和元年十一月場所

【力士の強さランキング (70 Entries)】



# 日々のモデル的中状況をモニタリング

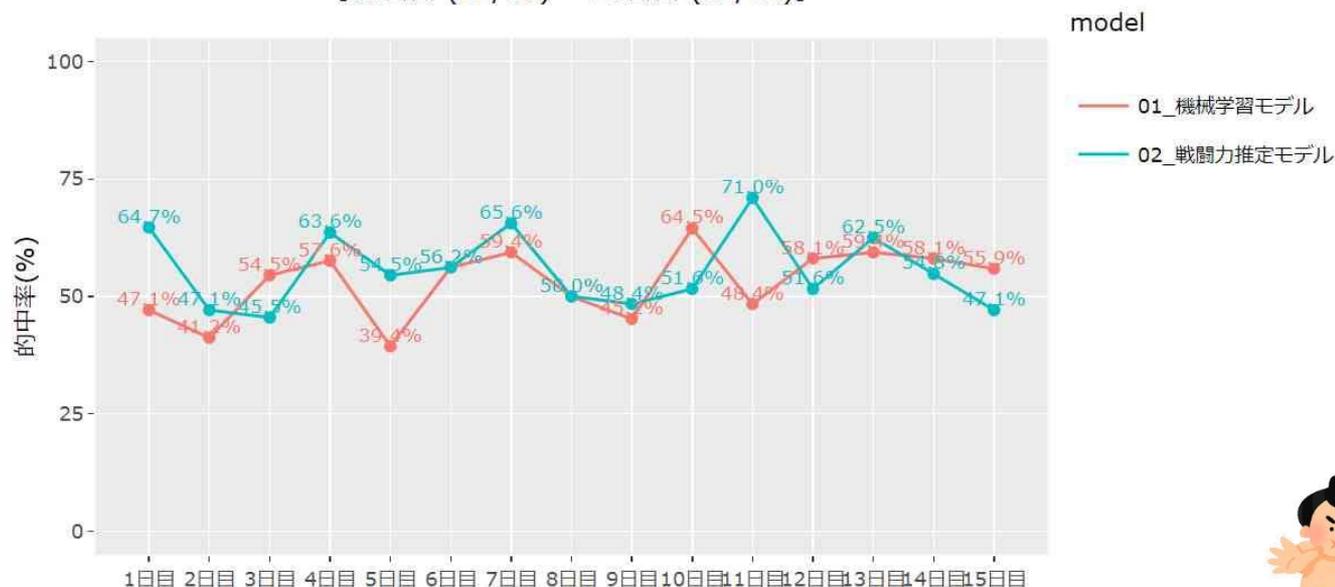
## 予測ツール概要(5/5): モデル別の中状況

番狂わせ、休場による不戦勝、反則負け等のイレギュラー時には的中率が低下する傾向

### 3. モデル別の中状況

令和元年十一月場所

【01日目 (11/10) ~ 15日目 (11/24)】



## <目次>

1. 分析の前提

2. 予測モデル

3. データでみる相撲

4. 今後の予定



# 立ち合いのスピードランキング (平成29年初場所/春場所)

## データでみる相撲(1/5):

白鵬  
1位



幕内力士 立ち合いのスピード (m/s)

1 白鵬	2.16	6 栃ノ心	1.96
2 日馬富士	2.11	7 荒鷲	1.94
3 琴奨菊	2.05	8 豪風	1.90
4 高安	2.03	9 逸ノ城	1.86
5 貴ノ岩	1.99	10 蒼国来	1.85

データ提供: Qconcept

立ち合いの速さも  
映像からデータ化  
される時代に!



# 立ち合いの腰の低さランキング (平成29年初場所/春場所)

データでみる相撲(2/5):

白鵬  
3位

幕内力士 立ち合いの腰の低さ (身長との相対値)

1 大栄翔	0.442	6 宝富士	0.46201
2 荒鷲	0.451	7 蒼国来	0.46204
3 白鵬	0.453	8 遠藤	0.468
4 宇良	0.458	9 隠岐の海	0.469
5 妙義龍	0.460	10 日馬富士	0.470

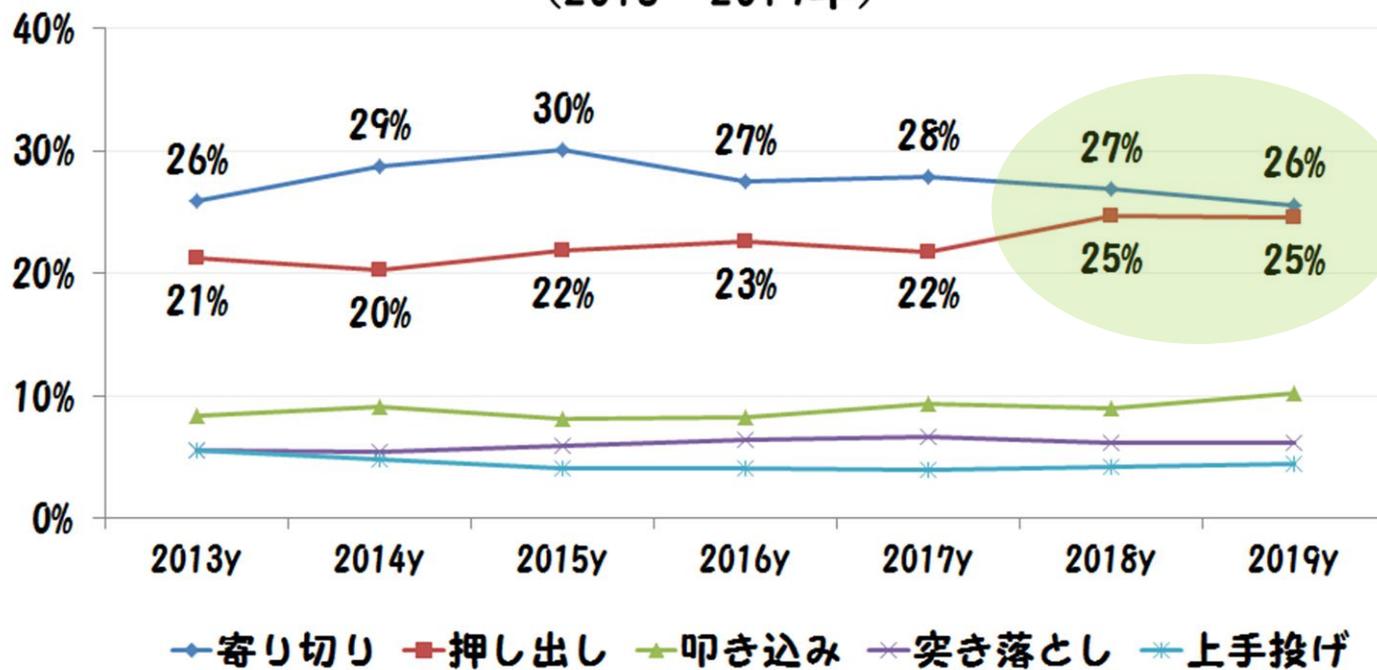


力士が撒く塩の量  
もデータで出して  
欲しいドスコイ!



# 近年は寄り切りがやや減少し、押し出しが増加 データでみる相撲(3/5):

決まり手上位5位の構成比推移  
(2013～2019年)



最近「押し相撲型の力士」が増えているらしいよ!



# 力士には、寄り切り型と押し出し型がいる模様

## データでみる相撲(4/5):

### ◆ 寄り切り型力士

No	力士	決まり手	決まり手数		構成比	
			寄り切り	押し出し	寄り切り	押し出し
1	琴奨菊	295	163	41	55%	14%
2	栃ノ心	276	144	13	52%	5%
3	朝乃山	136	62	20	46%	15%
4	逸ノ城	245	111	32	45%	13%
5	稀勢の里	269	120	61	45%	23%
6	照ノ富士	204	91	16	45%	8%
7	遠藤	282	115	44	41%	16%
8	宝富士	297	118	42	40%	14%

### ◆ 押し出し型力士

No	力士	決まり手	決まり手数		構成比	
			寄り切り	押し出し	寄り切り	押し出し
1	北勝富士	138	21	70	15%	51%
2	誉富士	237	16	115	7%	49%
3	貴景勝	138	2	66	1%	48%
4	玉鷲	300	15	143	5%	48%
5	阿武咲	211	23	100	11%	47%
6	琴勇輝	302	2	128	1%	42%
7	大栄翔	238	18	93	8%	39%
8	御嶽海	226	57	87	25%	38%

※ データ期間: 2013/01~2019/09

力士を得意な決まり手で分類すると面白そうだよ!

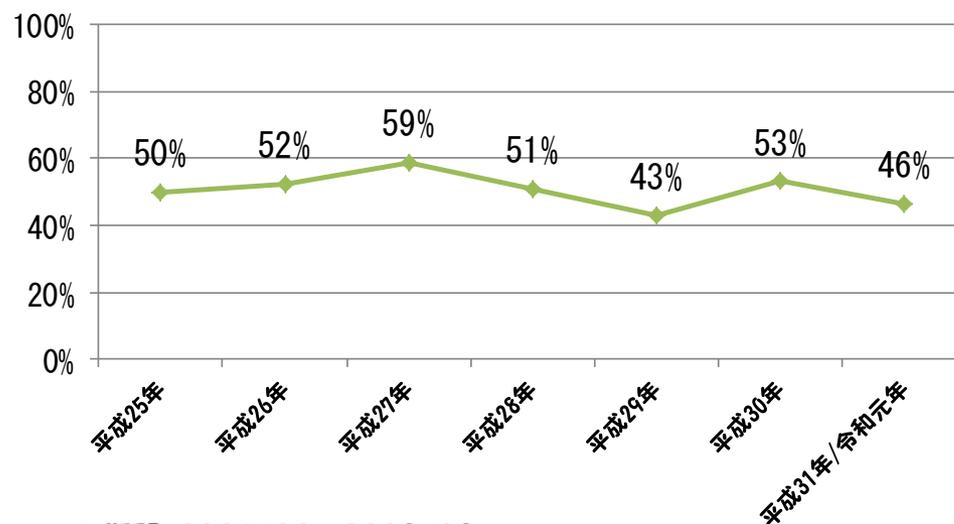


# 八百長って本当にあるの？

## データでみる相撲(5/5):

◆ 14日目までで7勝7敗の力士の千秋楽の勝率（幕内・十両のみ）

年	千秋楽に勝利した力士	14日目までで7勝7敗の力士	千秋楽の勝率
平成25年	36	72	50%
平成26年	34	65	52%
平成27年	40	68	59%
平成28年	30	59	51%
平成29年	24	56	43%
平成30年	35	66	53%
平成31年/令和元年	26	56	46%
計	225	442	51%



※ データ期間:2013/01~2019/09

幕内・十両については八百長は無さそうですね！



## <目次>

1. 分析の前提

2. 予測モデル

3. データでみる相撲

4. 今後の予定



# 馴染みの薄い相撲を楽しく観戦したい 今後取り組みたいこと:

## 1. 予測精度の向上

- ・ 今場所中の情報をデータに織り込む予定
- ・ ハイブリッドモデルの追加

## 2. 予測対象の拡大

- ・ 珍しい決まり手の予測
- ・ 荒れる日(=番狂わせが多い日)の予測
- ・ 今場所に躍進する力士の予測
- ・ 休場する力士の予測

## 3. 楽しく観戦する為の補助情報の活用

- ・ 力士の画像をスクレピングし、レポートに追加
- ・ 場所中の勝敗数ランキングの表示
- ・ 力士のSNS投稿(Twitter等)の活用



(おまけ)  
画面広々、一文字力士対決！！



<https://togetter.com/li/1113296>



*End of Document*

