

テキストマイニングを用いた競輪情報の分析

滝澤 みこ[†] 千田乃々佳[†] 大塚 真吾[†]

[†] 神奈川工科大学情報学部情報工学科 〒 243-0292 神奈川県厚木市下荻野 1030

E-mail: [†]{s1921155,s1921137}@cco.kanagawa-it.ac.jp, ††otsuka@ic.kanagawa-it.ac.jp

あらまし 近年、高速にテキスト分類を行うことができる自然言語処理ライブラリの fastText に注目が集まっており、様々な分野で利用されている。一方、スポーツ業界では AI を取り入れた取り組みが盛んになってきており、その中でも「勝敗予想」は、野球、サッカー、オートレース、競馬、競輪など多くの人気スポーツで用いられている。本研究では、その中でも競輪に着目し、レース予想を行うことを試みる。具体的には、各レースの情報を Web から収集し、レースごとに各選手の勝率や並び順などを車番順に並べたものにレース結果をタグ付けしたものを学習データとし、fastText で学習を行う。次に、タグ付がされていないレース情報をテストデータとしてレース結果を予想する。評価指標については、レース数に対して予想が的中した割合である的中率の他に、購入した金額に対して回収できた金額の割合を示す回収率を用いた。実験は 15,000 レースをトレーニングデータ、1,000 レースをテストデータとして行い、エポック数などのパラメータ調整を行うことで、回収率が向上することがわかった。

キーワード スポーツデータ活用, ゲーム・娯楽, テキストマイニング, 自然言語処理応用, fastText

1 はじめに

近年、スポーツ業界では AI による様々なソリューション技術が開発され注目を浴びている。それにともない、日本政府はスポーツ産業の市場規模を 2025 年後には 15 兆円にする目標が掲げられており、期待と同時に成長が見込まれている。スポーツ業界の AI によるソリューション技術の中でも「勝敗予想・予測」は、野球、サッカー、オートレース、競馬、競輪など多くの人気スポーツで用いられている。本稿で扱う競輪は日本発祥のスポーツであり、1948 年に初めて小倉レースが開催された。その後、1896 年にはオリンピックの正式種目として実施されている。競輪は公営競技の一つであり、観客は 1 着から 3 着までの選手を選手情報などから予想・予測する。現在、AI を用いて競輪予想を行っているサイトがいくつか存在しているが、投資した金額に対してそれ以上の金額が当たることが少ないのが現状である。

本稿では、過去の競輪レース情報とその結果を用いて機械学習を行い、未知のレースの予測を行うことを目指す。機械学習には Meta 社が公開している自然言語処理ライブラリである「fastText [1]」を用いる。本稿での「予想」とは、あるレースにおいて 1 着から 3 着までの選手を予測することを指す。また、レース数に対して的中した割合である的中率や、購入した金額に対して回収できた金額である回収率などについても高い精度を目指す。

2 研究の背景, および, 関連研究

近年、スポーツとテクノロジーを掛け合わせた言葉である「スポーツテック」に注目が集まっており、スポーツ選手のサポートや競技に関する環境改善や道具の開発などに用いられている。AI によるサッカーや競輪、競馬の勝敗予想・予測もス

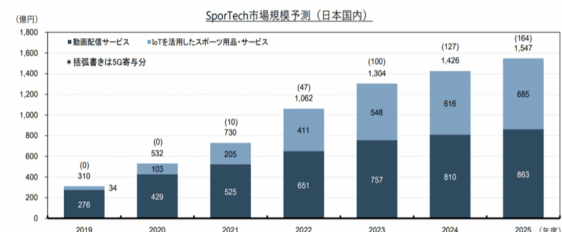


図 1 スポーツテック市場規模の予想

スポーツテックの事例の 1 つである。

図 1 は 2019 年度に発表したスポーツテックの市場規模を予想した調査結果である [2]。2019 年時点では 310 億円であるが、6 年後の 2025 年では約 5 倍の金額である 1,547 億円となり、急速な成長が予想されている。同時に、スポーツ業界での新たなソリューションの需要の高まりを読み取れることができる。次に、スポーツテックには重要な要素の 1 つに「創る」がある。これは、従来の定義を超えた新たなスポーツを創出するソリューションを開発し続けることを意味する。事例としては、水泳選手が使用する「FORM SwimGoggles」というゴーグルにスマートディスプレイを搭載し、データ収集を行うことができるサービスがあり、カナダ発のスタートアップ企業である FORM によって開発されている。このディスプレイは水泳選手の距離やタイムなどをリアルタイムに保存でき、選手やコーチが情報収集を行う労力を軽減することが可能である。

公営ギャンブルの予想に関しては、情報処理学会の学会誌において競馬予想に関する解説記事が掲載されるなど、注目が集まっている [3]。また、競馬を対象としては、時系列オッズデータを用いて競馬予想について優れた情報を持った投票者を窺い知することで、合理的な投資戦略を行う研究 [4] などがある。

本稿で対象としている競輪の予想に関しては、Web 上にあって多くのサイトが存在しているが、解析・分析内容が公開さ

れているケースは皆無である。関連研究としては、レース展開を含むレース結果の予測器を Transformer により構築し、予測器の出力に基づいて、事前に人手で設定したテンプレートを使用するルールベース手法により、予想記事の自動生成を行う研究 [5] がある。また、出走選手同士の対戦を考慮した「対戦モデル」を深層学習により構築し、「予想モデル」により連勝単式の予想を行う研究 [6] などもある。

競輪における AI の予想事例としては、[5] の研究グループと車券投票サイトの「チャリロト.com」が「AI 競輪」というサイトを運営している [7]。このサイトでは、競輪予想などを無料で公開しており、予想結果だけでなく展開予想のような一言コメントが生成されているという特徴がある。また、過去の成績について、予想数や回収率が通算と年月集計で公開されており、月間回収率の最高 123% となっている。

サイバーエージェントが出資している WINTICKET [8] では、サイト内の AI による予想の事を「AI 予想屋」と呼び、この予想が 1 着になる可能性が最も高い選手◎ (本命)、本命の次に 1 着に入る可能性が高い選手○ (対抗)、展開によっては 1 着に入る可能性がある選手△ (単穴)、1 着に入る可能性は低い展開によっては 2 着以内に入る可能性が高い選手× (連下) を予想し、選手名を記載している。

その他、Loto Place [9] では、クイックピック方式での購入が可能であり、競輪に詳しくない人でも宝くじ感覚で投票ができるのが特徴である。予想ページでは、過去 5 万分のレースのデータから AI による予想が記載されており、的中率を重視した予想エンジンである Type.A (アペレージ型) と的中率は低いと当たると高額配当になる Type.S (スラッガー型) の 2 つの AI エンジンで構成されている。

3 競輪の概要

日本においては、競輪・競馬・競艇・オートレースが公営ギャンブルとして許可されており、これらの競技は自治体や特定の公営団体が開催し、投票券を販売している。その中でも、競輪は自転車を用いて行われるスポーツであり、現在、国内に 43 箇所の競輪場が存在している。レースは通常 7 名、または、9 名で行われることが多く、最小人数は 5 名である。

男性の選手は実力に応じて SS 級から A3 級までの 6 クラスに分かれており、女性の選手は全て L1 級の 1 クラスとなっている。全てのレースは GP、GI、GII、GIII、FI、FII の 6 つのグレードに格付けされており、SS、S1、S2 級の選手は FI と呼ばれる S 級シリーズや GI、GII、GIII などグレードレースと呼ばれる獲得賞金の高いレースに参加する。A1、A2 級の選手は FI で開催される A 級レースや、A 級選手のみ (一部 L1 級レースも開催される。) で行われる FII と呼ばれるレースに参加する。最後に、A3 級の選手は FII で開催される A 級チャレンジレースに参加する。

FII 開催のレースは通常 7 名で行われ、それ以外のレースは通常 9 名で行われる¹。

1: コロナ禍の影響で 2020 年 7 月から 2023 年 1 月時点まで FI は通常 7 名

3.1 開催時間と開催期間

競輪は開催時間に応じて、4 つの形態がある。各開催の詳細について以下で述べる。本研究では、ミッドナイト競輪において、AI 予想を行うためレースに参加する選手数は 5-7 名となる。モーニング開催

1 つの開催は 3 日間で行われており、1 日に行われるレース数は 10 レース、または、7 レースとデイ開催より少ない。また、開催されているレースは FII のみであるため、1 レースに出場する選手数は 5-7 名である。

デイ開催

一般的な開催方式であり、1 つの開催は 3 日間から長いものでは 6 日間に渡って行われる。主に GI、GII、GIII、FI が行われる。ナイター開催

1 つの開催は 3 日間で行われており、主に FI と FII が開催されている。また、近年の競輪人気の影響で GI、GII、GIII も開催されることがある。1 日 12 レース程度のレース数である。

ミッドナイト開催

無観客で開催されるため、自宅などからインターネット環境を通じて投票や観戦を行う。他の開催時間帯の通常のレースは 5 周回に対して、ミッドナイトは 1 週少ない 4 周回でレースが行われ競争距離が短いという特徴がある。主に FII が開催されている。

3.2 賭式について

競輪では表 1 のように、7 種類の賭式が存在しており、競馬や競艇などにある単勝や複勝の販売は無い。また、2 枠単と 2 枠複は 7 車立てのレースでは発売されないため、主にモーニング開催やミッドナイト開催されている FII ではこれらの車券は販売されていない。

表 1 賭式の種類

賭式名	説明	9 車の確率 (7 車の確率)
2 枠複	枠番で着順にかかわらず 1-2 着を予想し、的中させる車券。	1/18(未発売)
2 枠単	枠番で 1・2 着を着順どおりに予想し、的中させる車券。	1/33(未発売)
2 車複	車番で着順にかかわらず 1-2 着を予想し、的中させる車券。	1/36(21)
2 車単	車番で 1・2 着を順番通りに予想し、的中させる車券。	1/72(42)
3 連複	車番で着順にかかわらず 1-3 着を予想し、的中させる車券。	1/84(35)
3 連単	車番で 1・2・3 着を順番通りに予想し、的中させる車券。	1/504(210)
ワイド	車番で 1-3 着内の 2 車を予想し、的中させる車券。 この車券のみの中車券は 3 つとなる。	3/36(3/21)

で行われている。また、L1 級の選手が参加するレースは全て 7 名である。

4 競輪データの収集

競輪のレース予測を行うためには、過去のレース情報を得ることが必要である。レース情報は予想サイトや投票サイトなどで公開されているため、Web サイトへの負荷に配慮しながら時間をかけて収集を行った。対象レースは、選手の力量を示す競走得点順に1 番車から並んでいるため比較的实力通りに決着がつくことが多く、また、7 車以下のレースで開催されている「ミッドナイト開催」に限定した。

今回は WINTICKET サイト内にある出走表と結果の Web ページから、表 2,3 に示す情報を抽出した。取得したレース情報は 2018 年 1 月から 2022 年 5 月までの約 4 年 4 ヶ月の間にミッドナイトで開催された 19,596 レースであった。取得した情報から以下を含むレースを取り除いた結果、16,105 レースとなったため、今回の実験では古いレースから 15,000 レースをトレーニングデータに、それ以降の 1,000 レースをテストデータとした。

- 新型コロナ等の発生により急遽中止になったレース。
- 同着が発生したレース。
- 出走表に【選手の並び順予想】が無いレース。

表 2 出走表データ

選手名	選手の名前.
登録府県	選手の登録府県.
級班	選手を实力順に 6 つのランク (SS・S1・S2・A1・A2・A3) に分けたもの.
脚質	選手が得意とする戦法 (脚のタイプ). 逃げ, 追込み, 両方の 3 種類.
期別	競輪学校の卒業期.
年齢	選手の年齢.
競争得点	選手の強さを得点化したもの. 選手個人の成績.
決まり手	1 着・2 着に入った際の決まり手 (逃・捲・差・マ) の回数.
B・H・S	B は最終バックストレッチ, H は最終ホーム ストレッチを先頭に通過した回数.S はスタート 回数でスタート直後に一番前を取った回数.
勝率 2 連勝率 3 連勝率	勝率は 1 着, 2 連勝率は 2 着, 3 連勝率 3 着 に入った確率を表す数値.
並び順予想	選手のコメントにもとづいたレース中の並び順.

表 3 結果表データ

レース開催地	レースの開催場所.(全国で 43 箇所)
周長 みなし直線 カント	レース開催地情報をもとに, 1 周の長さ, 直線の長さ, コーナーの角度情報を追記する.
レースグレード	A 級,A 級チャレンジ,I 級のどれか
レースクラス	予選, 準決勝, 一般など
着順	レースの着順.
レース結果 払戻金	賭け式ごとの結果と払戻金. (2 車複, 2 車単, 3 車複, 3 連単, ワイド)

5 取得データの傾向

前節で述べたようにミッドナイト開催では 1 番車から実力がある選手が並んでいるため、例えば、1=2 などの結果が出やすと考えられる。そこで、トレーニングデータ 15,000 レースとテストデータ 1,000 レースにおける各車券について出現頻度を調べた。各車券で当たり車券になる確率が高い車券上位 10 件を表 4 に示す。どの賭式においても、1 番車と 2 番車が当たる車券に絡むことが多く、組合せ数が比較的少ない 3 連単以外の賭式では、上位 10 件の合計が 5 割を超えていることがわかる。2 車複の場合は上位 2 点, 2 車単の場合は上位 4 点で 25%を超えており、これらの車券を買えば、4 回に 1 回は当たることがわかる。また、3 連単は各車番の順序も的中しなければならず、組合せ数も 7 車立ての場合は 210 点と多いことから、他の車券に比べて当たり車券になる確率が低いことがわかる。

表 4 各賭式における出目上位 10 件

		2車複(21通り)				2車単(42通り)				3連複(35通り)				3連単(210通り)	
出目 順位	出目	トレーニング		テスト		トレーニング		テスト		トレーニング		テスト		トレーニング	
		出現確率(%)	出目	出現確率(%)	出目	出現確率(%)	出目	出現確率(%)	出目	出現確率(%)	出目	出現確率(%)	出目	出現確率(%)	出目
1	1=2	17.95	1=2	28.10	1=2	11.44	1=2	19.50	1=2=3	11.21	1=2=3	20.60	1=2=3	3.59	1=2=3
2	1=3	10.98	1=3	14.70	1=3	7.02	1=3	10.10	1=2=4	8.69	1=2=4	12.50	1=2=4	2.79	1=2=4
3	1=4	9.28	1=4	11.10	2=1	6.51	2=1	8.60	1=2=5	6.76	1=2=5	7.50	1=2=5	2.34	1=3=2
4	1=5	7.17	2=3	7.30	1=4	5.77	1=4	7.70	1=3=4	5.65	1=3=4	7.20	1=3=2	2.29	2=1=3
5	2=3	6.22	1=5	7.10	1=5	4.54	1=5	4.60	1=3=5	5.06	1=3=5	5.30	2=1=3	1.97	1=4=2
6	1=7	6.19	1=6	5.30	3=1	3.96	3=1	4.60	1=4=5	4.35	1=2=6	5.20	1=4=2	1.67	1=2=5
7	2=4	5.34	2=4	5.00	4=1	3.51	3=2	4.50	1=2=7	4.27	2=3=4	4.00	2=1=4	1.83	1=3=4
8	2=5	4.83	2=5	3.40	1=7	3.40	1=6	4.30	1=2=6	3.79	1=3=6	3.80	1=3=4	1.61	1=4=3
9	3=5	4.15	3=4	3.00	2=3	3.37	4=1	3.40	1=4=7	3.52	1=4=5	3.60	1=2=6	1.60	2=1=5
10	3=4	3.79	3=5	2.50	2=4	3.12	2=4	2.90	2=3=4	3.41	1=2=7	3.40	1=3=5	1.40	1=2=7
合計		75.90		87.50		52.64		70.20		56.71		73.10		20.89	

次に車番と着順の関係について調べ結果を図 2 に示す。図の横軸は各車番における着順確率を示しており、例えば、1 着の一番高い点 (33%) は 1 番車の 1 着率を示している。また、2,3,4 番車の順で確率が高いことがわかる。一方、7 位 (最下位) になる確率が高い車番は、6 番車であり全体の 24%を占めている。次いで、7 番車が 19%を占めている。前節で述べたように、ミッドナイト開催では各選手の強さの指標となる競走得点が高い順に 1 番車から割り当てられており、競輪では車番が内側の選手がスタート直後に誘導員の後ろを確保しやすいなどが要因になっていると推測している。また、推測ではあるが、6 番車と 7 番車の傾向が入れ替わっている点については、毎年 1 月と 7 月に各選手の所属クラスの入替えが行われており、下のクラスから上がってきた選手は能力は高いものの競走得点が低いため 7 番車になってしまうことがある。さらに、7 月には新規の競輪選手 (ルーキー) の加入があり、ルーキーは競走得点が低い、潜在能力が高い選手も多いため、7 番車でも 1 位になることなどが少なからず影響していると考えている。

6 fastText を用いて当たり車券の予測

実験では fastText のテキスト分類機能を用いて当たり車券の予測を行った。fastText は本来は、1 つの文章に 1 つのラベルを付与したものを大量に集めたものをトレーニングデータとし、学習を行うことでモデルを生成する。その後、そのモデルに対してラベルの付いていない文書を与えるとラベルの推定を

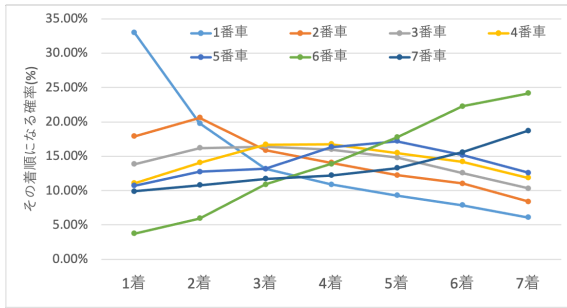


図 2 車番と着順の関係

行う。今回は、表 2,3 のデータ²を以下のように各値ごとにスペースで区切り、ラベルは当たり車券として学習を行いモデルの生成を行った。この例では当たり車券(3 連単)、レース場名、周長、みなし直線、バンクの角度、レースグレード、レースクラスと続き、その後は 1 番車から 7 番車までに選手情報が続く。最後に選手の並び順の予想となる³。

[_label_3-1-7 maebashi 335 46.7 36 A 級 特選 26 歳 92.00
4 1 追 0 0 4 2 2 4 4 13 8.6 26.0 43.4 29 歳 85.93 0 12 逃 5 5
1 0 7 4 3 16 23.3 36.6 46.6 42 歳 85.67 3 8 両 1 3 4 0 3 5 2
19 10.3 27.5 34.4 41 歳 89.50 0 0 追 0 0 1 3 1 3 0 20 4.1 16.6
16.6 46 歳 91.75 2 0 追 0 1 4 1 1 5 8 10 4.1 25.0 58.3 42 歳
82.57 3 0 追 0 0 3 2 0 5 3 18 0.0 19.2 30.7 37 歳 86.79 1 0 追
0 0 3 3 1 5 0 23 3.4 20.6 20.6 315/6/247]

モデル作成後は以下のようなラベルの付いていないデータと予測点数を与える。

[ogaki 400 56 30.38 A 級 特一般 32 歳 86.14 15 13 逃 5 3
2 0 3 7 3 15 10.7 35.6 46.4 49 歳 82.56 1 0 追 0 0 0 0 0 2
14 0.0 0.0 12.5 43 歳 82.27 15 0 追 0 0 4 1 4 1 4 20 13.7 17.2
31.0 45 歳 81.41 1 0 追 0 0 2 0 1 1 3 12 5.8 11.7 29.4 32 歳
80.50 0 0 追 0 0 0 2 0 2 3 19 0.0 8.3 20.8 32 歳 79.53 0 7 逃
2 3 4 0 6 3 2 15 23.0 34.5 42.3 43 歳 78.70 1 3 両 1 1 0 0 0 2
1 14 0.0 11.7 17.6 13/5/72/64]

例えば、予測点数を 24 点とすると、以下のような結果が返ってくる。ラベルの予測された当たり車券であり信頼度が高い順に並んでいる。数値は信頼度を示している。

((('_label_1-2-3', '_label_1-2-4', '_label_1-2-5', '_label_1-3-2', '_label_2-1-3', '_label_1-4-2', '_label_2-1-4', '_label_1-3-4', '_label_1-2-6', '_label_1-3-5', '_label_2-1-5', '_label_1-4-3', '_label_3-1-2', '_label_1-4-5', '_label_1-5-2', '_label_2-3-1', '_label_1-5-3', '_label_1-5-4', '_label_1-2-7', '_label_2-4-1', '_label_3-2-1', '_label_1-3-6', '_label_2-1-6', '_label_7-1-2'), array([0.03288206, 0.02691363, 0.0218959, 0.02161153, 0.01817421, 0.01620723, 0.01514563, 0.01511675, 0.01491531, 0.01328143, 0.01298939, 0.01239963, 0.01185606, 0.0117694, 0.01155519, 0.01147211, 0.01064028, 0.01059781, 0.01048894,

2：着順、レース結果、払戻金は除く

3：この例では 3,1,5 番車が並んで走り、その後に 6 番車が単騎で走り、2,4,7 番車が並んで走ることを意味している。

0.00971527,0.00870783, 0.00869684, 0.00860737, 0.00858748]))

fastText ではモデル作成時に様々なパラメータを設定することが可能であるため、各種パラメータを変更したが、epoch 数以外では大きな変化が見られなかったため、epoch 数に関する実験結果を示す。以降の実験結果は前節で述べた通りトレーニングレース数を 15,000 レース、テストレース数を 1,000 レースとしている。

実際の車券を購入する場合、1 点のみ購入することは稀であり、通常はオッズを見なら、3 点、6 点、12 点など複数の車券を購入する。今回の実験においても、1 点と 3 点から 3 点刻みで 24 点まで予想した場合的中率と回収率を調べた。的中率とは、予想し購入した買い目が的中する確率であり、予想点数が多いほどの中率は上がる。今回のテストデータは 1,000 レースのため、的中したレース数を 1,000 で割り 100 を掛けたものである。また、回収率とは、購入した金額に対して回収できた金額の割合を示す。今回の実験では、予想した車券は全て 100 円で購入したと仮定した場合の回収率としている。例えば、1 レースに 1 点予測した場合は 100 円となり、テストデータのレース数が 1,000 レースのため、100,000 円購入したときに戻ってくる払戻金の割合となる。同様に 1 レースに 24 点予測した場合は、2,400,000 円購入したときに戻ってくる払戻金の割合となる。

7 実験結果

各賭式における epoch 数と的中率・回収率の関係を図 3,4,5,6 に示す。縦軸は的中率と回収率であり、横軸は epoch 数である。なお、横軸は対数としている。1 点や 3 点などの点数は 1 レースあたりに予想する車券の数を示しており、例えば、24 点のグラフは 1 レースにつき車券を 24 通り予想していることを示している。予想点数が多いほどの中率が上昇するという特徴から、どの賭式においても、予想点数が多いほどの中率が良く、また、予想点数が少ない時は epoch 数が 1,000 を超えた辺りから変化が少なくなる傾向がある。

2 車複の結果である図 3 を見ると、的中率に関しては 21 点予想を除いて、epoch 数が大きくなるにつれて減少していることがわかる。回収率に関しては、他の点数と比較して 1 点予想が若干良いが、大きな違いは見られなかった。なお、2 車複の組合せは 21 点のため、21 点予想は全ての車券を予想していることから、的中率は 100%、回収率は 56.78%という値で一定となっている。2 車単の結果である図 4 を見ると、的中率に関しては epoch 数が大きくなるにつれて減少しているが、100 万回のときに若干上昇していることがわかる。回収率に関しては、他の点数と比較して 1 点予想が良く、epoch 数が 100 万のときに 90%を超える結果が得られた。

3 連複の結果である図 5 を見ると、的中率に関しては epoch 数が大きくなるにつれて減少している。回収率に関しては、1 点予想の変化が多いものの、大きな違いは見られなかった。3 連単の結果である図 6 を見ると、的中率に関しては epoch 数が大きくなるにつれて減少していき、1,000 を超えて辺りから、

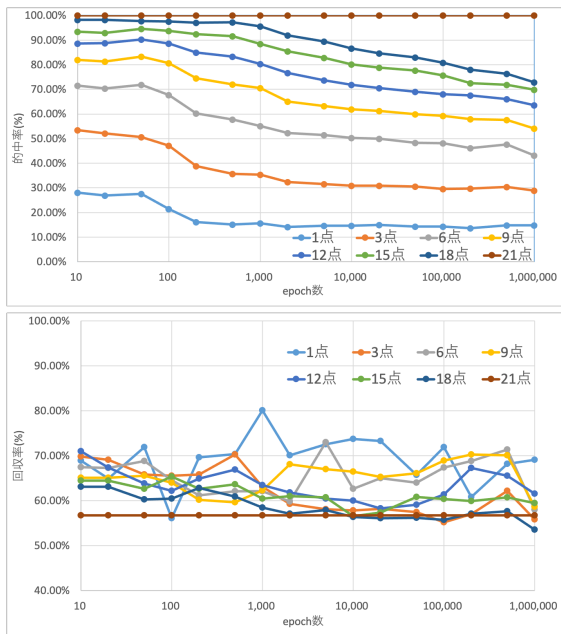


図 3 epoch 数と的中率・回収率 (2 車複)

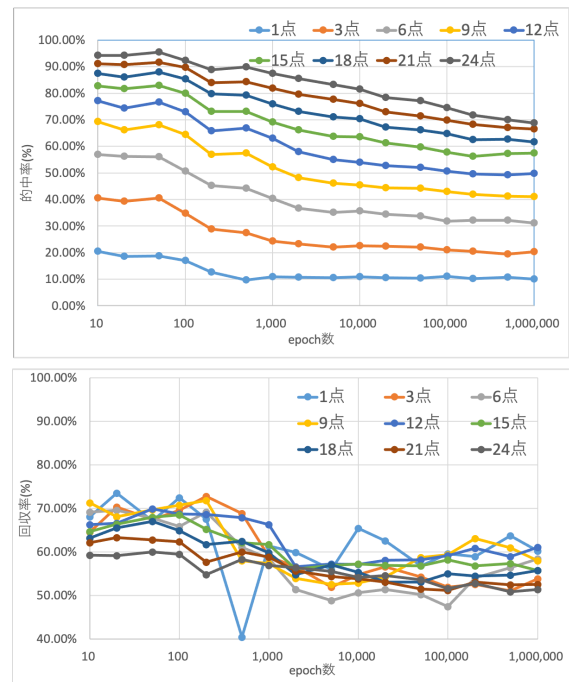


図 5 epoch 数と的中率・回収率 (3 連複)

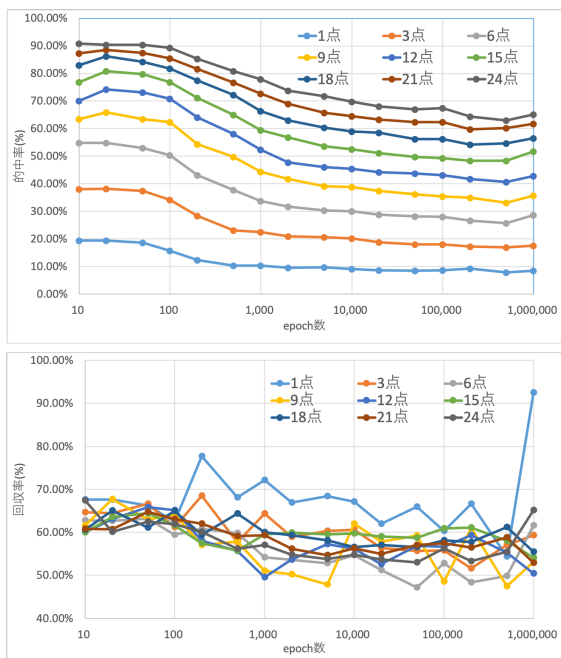


図 4 epoch 数と的中率・回収率 (2 車単)

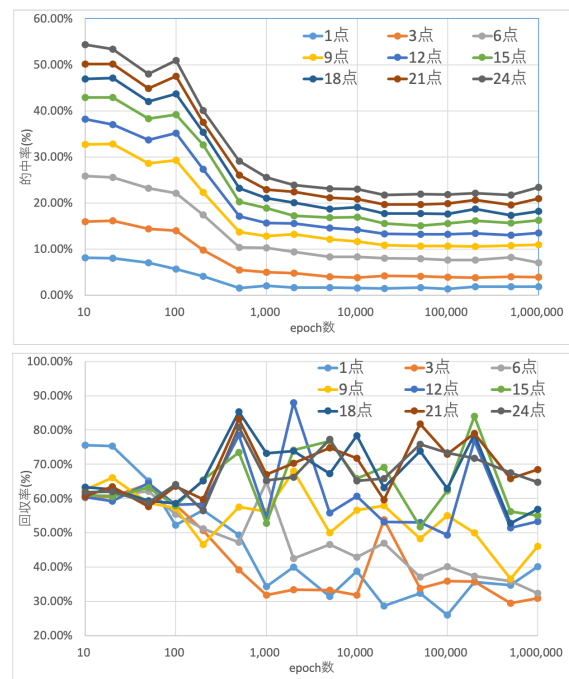


図 6 epoch 数と的中率・回収率 (3 連単)

変化が少なくなり 100 万の時に若干上昇していることがわかる。回収率に関しては、他の賭式と比べて変化が大きく、epoch 数が大きくなるにつれ、購入点数が多い方が回収率が良くなっていることがわかる。

次に、学習をどれだけ行っているのかを確認するために、epoch 数と 3 連単の予想車券の種類を調べた。その結果を図 7 に示す。ミッドナイト開催は 7 車で行われるため、3 連単の車券の組合せは 210 通りである。したがって、210 に近ければ様々な種類の車券を予想しているため、学習が適切に行われていると考えられる。一方、予想車券の種類が 1 に近いほど、毎レースで同じ買い目を予想しているため、学習効果は薄いと考

えられる。epoch 数が小さい場合は予想車券の種類が少ないため、単純に統計情報に依存して買い目を購入していると考えられ、表 4 で示した高頻度に出現する 1-2-3 や 1-2-4 などを用意している。図から、どの予想点数においても epoch 数が 500 を超えると、予想する買いの種類が多くなっており、1 点予想以外はほぼ全ての組合せを予想していることがわかる。他の賭式についても、ほぼ同様な結果が得られた。

最後に、トレーニングデータとテストデータの内容と的中率・回収率の関係について実験を行った。これまで、表 2,3 のデータの全てを用いて予想を行ったが、一部のデータのみを用いた

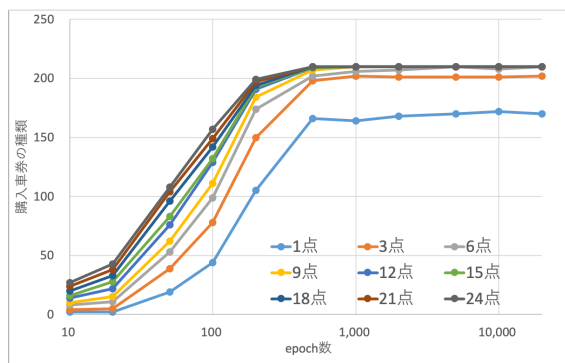


図 7 epoch 数と予想車券の種類 (3 連単)

ときの的中率と回収率について調べた。epoch 数を 100,000 とした結果を図 8 に示す。縦軸は的中率と回収率であり、横軸は予想点数を示す。また、「選手情報」は「開催場所特性」と比較するために、「勝率など」「1 着数など」「選手の特性 (決まり手)」「年齢と競走得点」を含んだものである。結果から、的中率に関しては、並び数や開催場所の特性、選手の特性や 1 着数などが良い一方で、全ての情報や選手情報は悪い傾向が見られた。一方、回収率を見ると、予想点数が多くなるにつれて、全ての情報や選手情報が良い結果となる傾向が見られた。開催場所の特性は的中率や回収率が比較的良好な傾向があることがわかる。その一方で、情報量が少ないと思われる選手の年齢と競走得点については良い結果を得ることができなかった。

これらの結果から、開催場所や選手の情報を利用すると、的中率は良いものの回収率はそれほど高くないことから、配当金が低い比較的中しやすいレースを的中させていると考えられる。一方、全て情報や選手情報については、的中率は低いものの予想点数が多くなるにつれて回収率が良いため、配当金が高めのレースを多めの予想点数で当てていると考えられる。開催場所特性については、的中率も回収率のどちらも比較的良好な結果が得られた。これは、選手によって開催場所の得意・不得意があるため、直近の成績が良く人気になる選手が特定の開催場所で悪い成績となったり、その逆に、直近の成績が悪く人気がない選手が特定の開催場所で良い成績となることが多々ある。この場合は的中すると高配当になる可能性が高いため、このような結果になったと推測している。

8 おわりに

本稿では、AI 予測を自然言語処理が可能な fastText を用いて、競輪の AI 予測を行った。実験結果から、回収率の向上には epoch 数や学習するデータの内容が関与していることがわかり、精度にばらつきがあるが、1 レースあたりの車券の購入数を調整することで、ある程度の回収率を得ることができた。今後は安定してより高い精度を出せるようにしていきたいと考える。実際には、トレーニングレース数、テストデータ数、入力する特徴量を増やすことで、回収率の向上を目指したい。

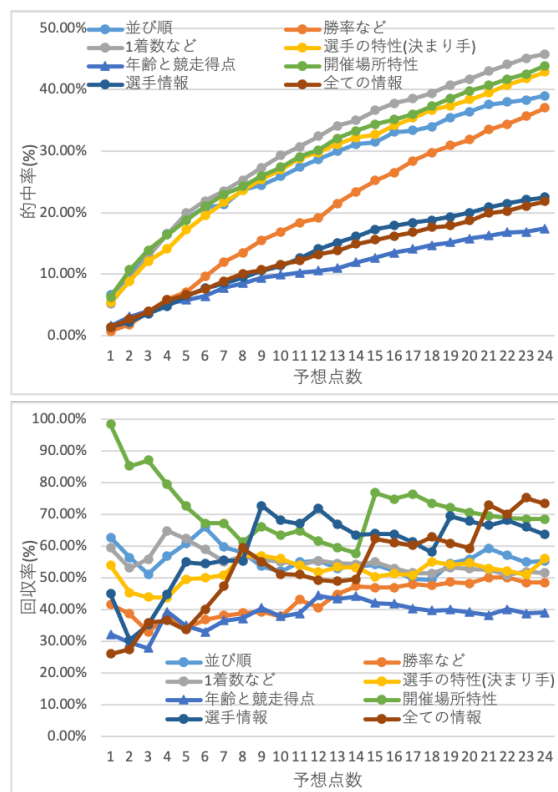


図 8 学習データの内容と的中率・回収率 (3 連単・epoch 数 100,000)

文 献

- [1] fastText, <https://fasttext.cc/>.
- [2] 「IT ナビゲーター 2020 年版」-Nomura Research Institute—SportTech (スポーツ) 市場, <https://www.nri.com/-/media/Corporate/jp/Files/PDF/knowledge/report/cc/mediaforum/2019/forum285.pdf>, 2019.
- [3] 寺沢 憲吾, 「情報学者が競馬予想に踏み出すときに知っておくべきこと」, 情報処理, Vol.60, No.2, pp.154–158, 2019.
- [4] 樋口 尚吾, 折原 良平, 清 雄一, 田原 康之, 大須賀 昭彦, 「投票行動時系列データ解析に基づく意思決定戦略に関する考察」, 信学技報, vol.115, No.97, AI2015-2, pp.7–11, 2015.
- [5] 吉田 拓海, 横山 想一郎, 山下 倫央, 川村 秀憲, 「競輪における予想記事の自動生成システムの開発」, 電子情報通信学会論文誌 D 情報・システム J105-D (1), pp.60–74, 2022.
- [6] 靱 勝彦, 「公営競技の予想モデルに関する研究」, 第 81 回全国大会講演論文集, Vol.2019, No.1, pp.103–104, 2019.
- [7] チャリ・LOTO, 北海道大学「AI 競輪」, <https://ai.chariloto.com/>.
- [8] WINTICKET(ウィンチケット), <https://www.winticket.jp/>.
- [9] LotoPlace(ロトプレイス), https://www.lotoplace.jp/statics/whats_ai_yosou.