

競輪予想記事の自動配信に向けた的中車券予測

吉田 拓海^{1,a)} 横山 想一郎¹ 山下 倫央¹ 川村 秀憲¹

概要：近年、競輪事業の活性化のために積極的な情報配信が望まれている。しかし、毎日 60 レースが開催され、レース前日に当日の出走者が決まるため、人手による予想記事の執筆の負担は大きい。そのため、予想記事を含むコンテンツの自動生成技術への高いニーズがある。本研究では、競輪の予想記事の自動生成に向けて、投票締切前の中車券の予測・配信を行うシステムを構築した。さらに、予測した的中車券に含まれる選手の戦績やライン構成を説明する記事の自動生成機能を実装した。

1. はじめに

スポーツ競技における記事の自動生成に関する研究は、盛んに行われている。野球においては、打者成績からインニング速報を生成する研究 [1] や、テキスト速報からインニングの要約文を生成する研究等 [2] が行われている。また、スポーツ以外の分野においても時系列数値データから概況テキストを自動生成する研究 [3] や、天気予報コメントを自動生成する研究 [4] 等が行われている。以上の研究は、どれもテキストを手で作成するコストを削減するために行われている。

本研究対象である競輪は、1 日平均約 60 レース実施されており、その各レースごとにレース結果を予想する記事が人手によって作成されている。また、当日のレースの出走者は前日のレース結果によって決定されるため、予想記事の作成は前日から当日までに行う必要がある。以上のことから、競輪の予想記事の作成にかかるコストは大きい。その一方で、競輪業界においては新規ユーザの獲得が大きな課題となっており、積極的な情報配信が求められている。そこで本研究では、競輪の予想記事を自動生成することを考える。近年、インターネット投票の売上が大幅に増加しているため、インターネット上での情報配信が競輪事業活性化に効果的であると考えられる。投票締切前の中車券の予測・配信を行うシステムを構築した。また、予測した的中車券に含まれる選手の戦績やライン構成を説明する記事の自動生成を実装した。

2. 競輪について

競輪とは、選手 9 人でバンクと呼ばれる競争路を周回し、

ゴールを競う日本発祥のトラックレースである。その他のトラックレースと異なる点として、ラインというものが挙げられる。ラインとは、レース中に選手が形成する縦列である。同じ県や、同じ地域の選手同士がラインを組むことが多く、同じレース内にラインを組む相手がいない選手は、単騎で走る。ラインの先頭になって走る先行選手は、走るペースや勝負を仕掛けるタイミング等を自由に組み立てることができるが、風の抵抗を一番に受けるため、体力を消耗する。一方で、先行選手の後ろで走る番手選手は、先行選手を風よけとして走ることができるため、先行選手よりも体力を消耗しない。その代わりに、他のラインに抜かれるないように、後続選手をブロックすることで先行選手を援護する。選手は最後のゴール前の直線に入るまで、ラインを組んでチームで走り、最後はライン関係なく 1 着を競う。このラインによって繰り広げられるレース展開は、その他の競技にはない競輪特有のものである。

競輪と同じ公営競技である競馬に関しては、ニューラルネットワークを用いて着順を予測する研究 [5] や SVM による予測 [6] 等、着順を予測する研究が数多く行われている。競輪と競馬との違いとして、ラインによる対戦者間の相互作用の強さが挙げられる。この相互作用を考慮することが競輪の予想において重要である。

3. 競輪における予想記事

現在、実施される全レースに対して人手により予想記事が作成され、競輪新聞や投票サイトにて配信されている。Web 上の投票サイトで配信されている既存記事 [7] の例を次に示す。

- 磯島が駆けて番手の野木が本命。鋭さ光るのは丸山だ。逆転の捲りに一考。目標の高鍋次第で大久保、攻め多彩な吉田も怖い。

¹ 北海道大学大学院情報科学研究科

^{a)} yoshida@complex.ist.hokudai.ac.jp

- 持ち前のスピードを全開なら⑤松本 ②蓮井でワンツー．自在③田中 ⑨米原の捲りや同県⑥山田にマークする④竹内や⑦内村も単級
- 藤縄一中澤の近畿勢に期待．隅田一高津，中部勢の抵抗も見ものだろう．
- 函館最終戦のオープニングレースは2分の争いとなった．ラインの先頭を走る自力型の先行力は，ほぼ互角とみるが109期のルーキー神山が主導権を握ると見て番手を回る①田中が鋭く追い込む．ラインの⑤松本が続き本線を形成．対し大阪勢も侮れない．ライン4車の利を活かし④斎藤が巧く駆けると番手を回る②陶器とのワンツーも狙える．

既存記事が含む情報として次の4つが挙げられる．

- 上位予想選手
- 上位予想選手に関する情報
- ラインに関する情報
- レース展開に関する情報

また，既存記事には次の2つの特徴が挙げられる．

- ライン決着の予想
- 複数パターンの展開予想

ライン決着の予想が既存記事の特徴となっている理由として，ラインによる選手間の相互作用がレース結果，レース展開に大きな影響を与えていることが理由として考えられる．

4. 記事生成アプローチ

既存記事に含まれる4つの情報の内，レース展開に関する情報については現時点では取得する手段がないため，本研究では次の3つの情報を含む記事を生成する．

- 上位予想選手
- 上位予想選手に関する情報
- ラインに関する情報

機械学習によつて的中車券を予測し，予測した的中車券に含まれる選手を上位予想選手として選択する．記事は予め用意したテンプレートに対し上位予想選手や，上位予想選手の過去成績，上位予想選手のライン構成などの条件に応じたテキストを当てはめることによって生成する．

5. 車券予測

予想記事生成のために，的中車券を機械学習によつて予測する．本研究では，選手2人のどちらが上位になるかを予測する二値分類器によつて，的中車券の予測を行う．本章では，二値分類器の出力からの中車券を予測する，予測モデルについて詳細を示し，学習・テストの結果を示す．また，入力特徴量の設定に関する比較実験の結果，予測モデルの回収率の評価について示す．以降では，次の条件を満たすレースを学習・テストの対象とする．

- 男性レース

- 競り無し
- 同着無し
- 欠損データの無いレース
 - (過去 n ヶ月競争得点を入力に用いる場合，過去 n ヶ月レースに出場して無い等の理由により競争得点が取得できない選手が存在するレースを除外する)

5.1 予測モデル

レースの情報と選手2人の情報を二値分類器に入力し，その出力値を1レース分集計する．集計した数値から車券的中確率を近似的に計算し，的中確率が最大のものを的中車券予測とする．以降では，二値分類器の入出力と車券的中確率の近似計算についてそれぞれ詳細を説明する．

5.1.1 二値分類器の入出力

二値分類器の入力としてレース情報と車番 i ，車番 j の選手情報を用い， $x_{i,j}$ と表記する．入力 $x_{i,j}$ の教師ラベル $t_{i,j}$ を次のように定義する． $place_i$ は車番 i の順位とする．

$$t_{i,j} = \begin{cases} 0 & place_i > place_j \\ 1 & place_i < place_j \end{cases} \quad (1)$$

入力を $x_{i,j}$ とした時の二値分類器の出力を $y_{i,j}$ と表記し， $y_{i,j}$ を車番 i が車番 j より上位になる確率として扱う．

5.1.2 車券予測の生成

二値分類器の出力値から車券予測を生成する．二値分類器の出力値は $y_{i,j} \neq 1 - y_{j,i}$ となっているため，出力集計時に式2のような補正を行う．

$$y_{i,j}(\text{補正後}) = \frac{y_{i,j} + (1 - y_{j,i})}{2} \quad (2)$$

式2の補正によつて $y_{i,j} = 1 - y_{j,i}$ が成立する．補正後の出力値を用いて車券的中確率を近似的に計算する．次の3つの確率をそれぞれ計算し，その積によつて車券的中確率を近似的に計算する．

- 車番 a が1着になる確率
- 車番 a が1着の時，車番 b が2着になる確率
- 車番 a が1着，車番 b が2着の時，車番 c が3着になる確率

各確率の計算式を次に示す．

$$p(1 \text{ 着} = a) = \frac{\prod_{j \neq a} y_{a,j}}{\sum_i (\prod_{j \neq i} y_{i,j})} \quad (3)$$

$$p(2 \text{ 着} = b | 1 \text{ 着} = a) = \frac{\prod_{j \neq a,b} y_{a,j}}{\sum_i (\prod_{j \neq a,i} y_{i,j})} \quad (4)$$

$$p(3 \text{ 着} = c | 1 \text{ 着} = a, 2 \text{ 着} = b) = \frac{\prod_{j \neq a,b,c} y_{a,j}}{\sum_i (\prod_{j \neq a,b,i} y_{i,j})} \quad (5)$$

式3,4,5の積を車番三連勝単式(三連単) $a-b-c$ の的中確率として以降扱う．的中確率の最も高い車券を車券予測とする．

5.2 入力特徴量

車券予測に使用した特徴量を表 1 に示す。バンク，単騎か否か，先行選手か否か，同じラインか否か以外の特徴量は平均 0 分散 1 に標準化する。

| レース情報 | 先行選手の平均競争得点の最大値 バンク |
|--------|--|
| 選手情報 | 車番 年齢 ギヤ倍率 平均競争得点 先行選手の平均競争得点 単騎か否か ラインの長さ(人数) 先行選手か否か 番手(ライン内での自身の位置) |
| 選手比較情報 | 同じラインか否か |

二値分類器の入力には，選手 2 人分の情報を使用するため，入力特徴量は表 2 のようになる。選手比較情報の同じラインか否かは，選手 1 と選手 2 が同じラインの場合は 1，異なるラインの場合は 0 となる特徴量である。

| レース情報 | 先行選手の平均競争得点の最大値 バンク |
|---------|--|
| 選手 1 情報 | 車番 年齢 ギヤ倍率 平均競争得点 先行選手の平均競争得点 単騎か否か ラインの長さ(人数) 先行選手か否か 番手(ライン内での自身の位置) |
| 選手 2 情報 | 車番 年齢 ギヤ倍率 平均競争得点 先行選手の平均競争得点 単騎か否か ラインの長さ(人数) 先行選手か否か 番手(ライン内での自身の位置) |
| 選手比較情報 | 同じラインか否か |

5.3 学習・テスト

学習，テストデータとして 2013 年 6 月 1 日 2016 年 11 月 1 日に実施された 64,232 レースを使用する。機械学習手法は予備実験の結果からニューラルネットワークを選択した。64,232 レースの前半 32,116 レースを使用してハイ

パーパラメータ探索を行う。ハイパーパラメータ探索はグリッドサーチによって行い，性能検証は 5 分割交差検証によって行う。分類の精度 (accuracy) が最良のハイパーパラメータセットを最適なものとして選択する。ハイパーパラメータの探索範囲を表 3 に示す。探索によって選択されたハイパーパラメータを次に示す。

● ハイパーパラメータ

– 中間層 : [256]

決定したハイパーパラメータを用いて，64,232 レースの後半の 32,116 レースに対し性能テストを行う。性能テストは 5 分割交差検証によって行う。評価項目として，上位 3 着の着順の中率，Top-K-accuracy を用いる。ここで，Top-K-accuracy はある順位の選手を上位 K 着以内に予測できたレースの割合とする。top3(1-2 着) は 1 着 2 着の選手を上位 3 着以内に予測できたレースの割合を表す。交差検証の結果の平均を表 4 に示す。予測精度の比較として，オッズを用いる。表 4 では最もオッズの低い車番三連勝単式を選択した場合の精度を示している。

表 3 ニューラルネットワーク：ハイパーパラメータ探索範囲

| ハイパーパラメータ | 探索範囲 |
|------------|---|
| 活性化関数 | ReLU |
| 出力層 | sigmoid |
| バッチサイズ | 256 |
| 学習率(初期値) | 0.001 |
| ドロップアウト率 | 0.5 |
| 学習 epoch 数 | 20 |
| 最適化手法 | Adam[8] |
| 中間層 | [2 ⁿ], [2 ⁿ , 2 ⁿ] (n=6,7,8) |

表 4 テスト結果

| 三連単の中率 [%] | | | |
|------------|----------------|----------------|--------------|
| 予測 | 4.6 | | |
| オッズ | 8.8 | | |
| | top1(1 着)[%] | top2(1 着)[%] | top3(1 着)[%] |
| 予測 | 38.4 | 58.7 | 71.7 |
| オッズ | 42.4 | 58.4 | 68.3 |
| | top2(1-2 着)[%] | top3(1-2 着)[%] | |
| 予測 | 24.0 | 42.2 | |
| オッズ | 30.3 | 43.5 | |
| | top3(1-3 着)[%] | | |
| 予測 | 13.6 | | |
| オッズ | 18.8 | | |

5.4 実験

入力特徴量の設定の違いによる予測性能の変化を調査する。

5.4.1 先行選手情報について

競輪はラインによる選手間の相互作用の影響が大きい。そのため，ラインに関する情報が重要となる。特に，ラインの先頭を走る先行選手の情報は重要である。本項では，先行選手に関する特徴量の設定について調査を行う。

5.4.1.1 実験設定

5.3 節では、入力特徴量として先行選手の平均競争得点を用いた。本項では入力特徴量として、先行選手に関する特徴量を全て用いた場合の予測性能を調査する。この時、入力特徴量は表 5 のようになる。

表 5 二値分類器の入力特徴量

| レース情報 | 先行選手の平均競争得点の最大値 バンク |
|------------|---|
| 選手 1 情報 | 車番 年齢 ギヤ倍率 平均競争得点 単騎か否か ラインの長さ(人数) 先行選手か否か 番手(ライン内での自身の位置) |
| 選手 1 の先行情報 | 車番 年齢 ギヤ倍率 平均競争得点 単騎か否か 先行選手か否か 番手(ライン内での自身の位置) |
| 選手 2 情報 | 車番 年齢 ギヤ倍率 平均競争得点 単騎か否か ラインの長さ(人数) 先行選手か否か 番手(ライン内での自身の位置) |
| 選手 2 の先行情報 | 車番 年齢 ギヤ倍率 平均競争得点 単騎か否か 先行選手か否か 番手(ライン内での自身の位置) |
| 選手比較情報 | 同じラインか否か |

学習・テストは 5.3 節と同様に行う。探索によって選択されたハイパーパラメータを次に示す。

- ハイパーパラメータ
 - 中間層 : [256]

5.4.1.2 結果・考察

テスト結果を表 6 に示す。結果としては、性能の大きな変化は見られなかった。このことから、先行選手に関する特徴量として、現時点では平均競争得点が予測性能に大きく関わっているということがわかる。今後の課題としては、平均競争得点以外の有効な特徴量を選択することが、予測性能の向上のために重要である。予測性能の低下が確認できないことや、今後の特徴量探索での特徴量の組み合

わせによる性能向上の可能性を考慮し、以降の実験では先行選手に関する特徴量を全て用いる予測を行う。

表 6 先行選手情報の比較

| 的中率 [%] | | | |
|---------|----------------|----------------|--------------|
| 競争得点のみ | 4.6 | | |
| 全て | 4.7 | | |
| | top1(1 着)[%] | top2(1 着)[%] | top3(1 着)[%] |
| 競争得点のみ | 38.4 | 58.7 | 71.7 |
| 全て | 38.4 | 58.8 | 72.0 |
| | top2(1-2 着)[%] | top3(1-2 着)[%] | |
| 競争得点のみ | 24.0 | 42.2 | |
| 全て | 24.0 | 42.5 | |
| | top3(1-3 着)[%] | | |
| 競争得点のみ | 13.6 | | |
| 全て | 13.8 | | |

5.4.2 特徴量の追加

機械学習において、入力特徴量の選択は重要である。有効な特徴量を使用することによって、予測性能の向上が期待できる。予測に有効であると考えられる入力特徴量を幾つか追加した場合の予測性能を調査する。

5.4.2.1 実験設定

5.3 節では、選手情報として表 1 で示したものを使用した。本項では、表 1 で示した特徴量に加えて表 7 に示す特徴量を加えた場合の予測性能を調査する。逃げ、捲り、差し、マークはレースの決まり手である。バック取り回数とは、ゴールラインの半周前にあるバックストレッチラインを先頭で通過した回数である。学習・テストは 5.3 節と同様に行う。探索によって選択されたハイパーパラメータを次に示す。

- ハイパーパラメータ
 - 中間層 : [256]

表 7 追加特徴量

| 選手情報 | 追加特徴量 |
|------|-----------------|
| | 過去 4 ヶ月の 1 着回数 |
| | 過去 4 ヶ月の 2 着回数 |
| | 過去 4 ヶ月の 3 着回数 |
| | 過去 4 ヶ月の 1 着率 |
| | 過去 4 ヶ月の 2 連対率 |
| | 過去 4 ヶ月の逃げ回数 |
| | 過去 4 ヶ月の捲り回数 |
| | 過去 4 ヶ月の差し回数 |
| | 過去 4 ヶ月のマーク回数 |
| | 過去 4 ヶ月のバック取り回数 |

5.4.2.2 結果・考察

テスト結果を表 8 に示す。結果から、特徴量の追加によって予測性能の向上が確認できる。有効な特徴量の探索は、今後の課題である。

表 8 特徴量追加の比較

| 的中率 [%] | | | |
|---------|----------------|----------------|--------------|
| 追加前 | 4.7 | | |
| 追加後 | 4.9 | | |
| | top1(1着) [%] | top2(1着) [%] | top3(1着) [%] |
| 追加前 | 38.4 | 58.8 | 72.0 |
| 追加後 | 40.4 | 60.7 | 73.1 |
| | top2(1-2着) [%] | top3(1-2着) [%] | |
| 追加前 | 24.0 | 42.5 | |
| 追加後 | 25.2 | 43.4 | |
| | top3(1-3着) [%] | | |
| 追加前 | 13.8 | | |
| 追加後 | 14.3 | | |

5.4.3 オッズの使用

5.3節の予測性能はオッズの予測性能よりも低い事が、表4からわかる。オッズよりも良い予測性能を示す予測器を生成するために、オッズを入力に使用するという方法が考えられる。本項では、表7の特徴量に加え、オッズを入力特徴量に追加した場合の予測性能を調査する。本実験では、車番二連勝単式車券のオッズを市場の予想する勝率へ変換し、その勝率を入力特徴量として使用する。車番 i の勝率を p_i 、車番二連勝単式 ($i-j$) のオッズを $odds_{i-j}$ として、式6によってオッズを勝率に変換する。

$$p_i = 0.75 \times \sum_{j, j \neq i} \frac{1}{odds_{i-j}} \quad (6)$$

学習・テストは5.3節と同様に行う。探索によって選択されたハイパーパラメータを次に示す。

- ハイパーパラメータ

- 中間層 : [128]

5.4.3.1 結果・考察

結果を表9に示す。オッズを入力に使用することで性能向上が確認された。特に、1着に関係する項目については、オッズの使用により性能の向上が確認でき、オッズの予測性能と同等かそれ以上の性能を示している。このことから、オッズの入力特徴量としての有効性が確認できる。一方で2着、3着に関係する項目については、性能の向上は確認できず、オッズの予測性能よりも劣っている。これは、車番二連勝単式のオッズを市場の予測する勝率へと変換したことによって、1着以降の順位に関する情報が欠損したことが原因であると考えられる。よって、同様の変換を2着、3着についても行い、その値を入力に使用することで、さらなる性能向上が期待できる。また、本実験では投票締切後の最終オッズを予測に使用したが、実レースの予測を行う場合、最終オッズは使用不可能であるため、最終オッズとは異なる投票締切前のオッズを予測に使用した場合の性能を調査する必要がある。これは今後の課題とする。

表 9 オッズを使用した結果の比較

| 的中率 [%] | | | |
|---------|----------------|----------------|--------------|
| オッズ未使用 | 4.9 | | |
| オッズ使用 | 5.0 | | |
| オッズ | 8.8 | | |
| | top1(1着) [%] | top2(1着) [%] | top3(1着) [%] |
| オッズ未使用 | 42.4 | 60.7 | 73.1 |
| オッズ使用 | 42.7 | 62.8 | 75.4 |
| オッズ | 42.4 | 58.4 | 68.3 |
| | top2(1-2着) [%] | top3(1-2着) [%] | |
| オッズ未使用 | 25.2 | 43.4 | |
| オッズ使用 | 26.1 | 45.0 | |
| オッズ | 30.3 | 43.5 | |
| | top3(1-3着) [%] | | |
| オッズ未使用 | 14.3 | | |
| オッズ使用 | 14.1 | | |
| オッズ | 18.8 | | |

5.5 回収率による評価

5.4節では、予測車券についての的中精度を評価した。本節では、回収率の評価を行う。

5.5.1 設定

本節では、回収率の評価を行うものとし、オッズが最小の車券の選択（以下、オッズ最小選択）と予測器の計算する発生確率が最大の着順に対応する車券の選択（以下、通常車券選択）、発生確率とオッズの積で計算される配当期待値による車券の選択（以下、期待値による車券選択）の回収率について示す。期待値による車券選択は、全車券の配当期待値を計算し、配当期待値がある閾値を超える車券を全て選択するという車券選択方法である。以降に示す回収率評価について、対象とする予測器はオッズを入力に使用したモノであり、予測器の学習・テストは5.3節と同様に行う。ここでの回収率は、選択車券に対し一定額を賭けた場合の回収率とし、期待値による車券選択については閾値6の結果を示す。

5.5.2 結果・考察

結果を表10に示す。この結果から期待値による車券選択によって回収率の向上が可能であることが分かる。予測器による予測は、車券的中精度についてはオッズの予測性能に劣っているが、回収率については予測器の出力値を用いた車券選択によってオッズを上回る性能を示す。

計算された期待値が正しい値であるならば、期待値1以上の車券に賭けた場合、理論上は回収率は100%を超える。しかし、そのような結果は得られなかった。予測器の計算する着順の発生確率が、真の値と異なっていることが原因であると考えられる。つまり、選手2人を比較した時の一方の選手の勝率として扱っている二値分類器の出力値が真の値と異なっているということである。回収率向上のために、着順の発生確率を真の値に近づけることが今後の課題である。

表 10 回収率の評価

| | 二車単 [%] | 三連単 [%] |
|--------------------|---------|---------|
| オッズ最小選択 | 76.5 | 79.3 |
| 通常車券選択 | 78.8 | 70.8 |
| 期待値による車券選択 (閾値: 6) | 101.6 | 92.0 |

6. 予想記事の生成

第 5 章で生成した車券予測に基づき予想記事を生成する。

6.1 記事生成の流れ

記事生成システムの概要を図 6.1 に示す。生成された車券予測と選手情報から文テンプレートに当てはめる修飾文を選択し、選択された修飾文を文テンプレートに当てはめることによって、予想記事を生成する。以降では、文テンプレート、修飾文の選択について詳細を説明する。

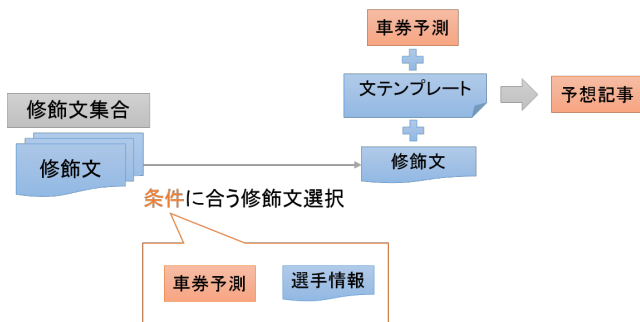


図 1 予想記事生成の概要

6.1.1 文テンプレートの設定

本研究で設定した文テンプレートを次に示す。

- ・ < 修飾文 1 > < 選手名 > が < 修飾文 2 > < 予測順位 > 着 < 修飾文 3 >

この文テンプレートに任意の文字列を当てはめることによって予想記事を生成する。文テンプレートに実際に文字列を当てはめると次のようになる。

- ・ 前日は 9 着だった②大山が自力で決めて 1 着。
 - < 修飾文 1 > : 前日は 9 着だった
 - < 選手名 > : ②大山
 - < 修飾文 2 > : 自力で決めて
 - < 予測順位 > : 1
 - < 修飾文 3 > : .

このように、文テンプレートに対して、説明したい状況に応じた < 修飾文 > を当てはめることによって、その状況を説明する文が生成可能である。

6.1.2 車券予測に基づく修飾文の選択

車券予測のライン構成に基づき修飾文を選択することにより、車券予測のライン構成を説明する予想記事を生成する。ライン構成は、1,2,3 着の選手の所属するラインと 1,2,3 着の選手のラインでの位置 (単騎, 先行, 番手) によ

て決定され、合計で 58 通り存在する。

6.1.3 選手情報に基づく修飾文の選択

選手に関する情報に基づき修飾文を選択することによって、選手個人に注目した記述をする予想記事を生成する。本研究では、以下のデータを選手情報として出走表から抽出した。

- ・ 過去 4 ヶ月の 1 着回数
- ・ 過去 4 ヶ月の勝率
- ・ 過去 4 ヶ月の 2 連対率
- ・ 年齢

抽出したデータに関して条件を設定し、その条件に応じて修飾文を選択する。本研究では、以下の条件を設定した。

- ・ 過去 4 ヶ月の 1 着回数 ≥ 10
- ・ 過去 4 ヶ月の勝率 ≥ 0.2
- ・ 過去 4 ヶ月の 2 連対率 ≤ 0.4
- ・ 年齢 ≤ 25
- ・ 年齢 ≥ 40

6.2 生成記事結果・考察

実際に生成された記事の例と既存記事 [7] を表 11,12,13 に示す。4 章で示した次の 3 つの情報を含む記事が生成されていることがわかる。

- ・ 上位予想選手
- ・ 上位予想選手に関する情報
- ・ ラインに関する情報

ラインに関する情報については、車券予測のライン構成に応じて異なる説明文が生成されていることが確認できる。選手個人に関する情報については、設定した条件に適合するレースには選手個人に関する記述が生成されている。しかし、今回の設定では選手個人に関する情報の記述が生成されないレースも存在した。選手個人に関する情報の記述が全てのレースにおいて生成されるような条件を設定することが、今後の課題として挙げられる。また、今回は生成の対象外とした、レース展開に関する記述も今後の課題である。

表 11 生成された記事例 1

| 2018 年 5 月 23 日向日町競輪場第 9 レース | |
|------------------------------|--|
| 既存記事 | 意地見せる地元高久保、器用な月森相手だが、松川が一枚上手か。 |
| 生成記事 | 勝率 22.2%の③松川が自力で 1 着。同ラインの⑤田中が③松川に続いて 2 着。③松川ラインで連携してトップ 2 独占。②月森ラインから⑨三宅が 3 着と予測。 |

表 12 生成された記事 2

2018 年 5 月 23 日岐阜競輪場第 4 レース

| | |
|------|---|
| 既存記事 | 二日目余裕で逃げ切った⑤山本のパワーが断トツ。ここも力任せの先行で豪快に逃げ切る。近畿勢⑦辻本が喰い下がると見るも、③安藤や①山中も気力次第か。 |
| 生成記事 | 勝率 33.3%の⑤山本が自分の脚で 1 着。同ラインの⑦辻本が⑤山本に続いて 2 着。②谷が 3 着。⑤山本ラインが別線を抑えて上位独占と予測。 |

表 13 生成された記事例 3

2018 年 5 月 23 日佐世保競輪場第 7 レース

| | |
|------|--|
| 既存記事 | 互角戦だが⑦滝本の自力に期待。①戸伏の逆転も。④内藤の連入十分。②飯塚も自力で単も。穴は⑥松尾の抜け出しから。 |
| 生成記事 | 二連対率 45.8%の⑦滝本が自力で 1 着。同ラインの番手①戸伏が滝本に続いて 2 着。⑦滝本ラインでトップ 2 独占。別線の⑥松尾が 3 着と予測。 |

7. 予測・配信システム

本研究で構築した、投票締め切り前の中車券の予測・配信を行うシステムについて説明する。予測・配信の流れを次に示す。

- (1) AM7:00 にレースの出走表データを取得する。
- (2) 取得した出走表データからレースの投票締め切り時間を取得する。
- (3) 取得した投票締め切り時間の 10 分前にオッズデータを取得し、出走表データからの中車券を予測する。
- (4) 車券予測と出走表データから予想記事を生成する。
- (5) 車券予測と予想記事をチャットサービスを利用して配信する。

出走表データとオッズデータは Web サイトから取得する。現在、本システムによって実レースに対して車券予測、予想記事生成を行っている。投票締め切り 10 分前のオッズデータは、実レース予測にオッズを入力特徴量として使用した際の予測性能を調査するために、蓄積中である。

8. まとめと今後の展望

本研究では、機械学習により競輪的中車券を予測し、予測結果に基づく予想記事の自動生成を行った。車券予測については、選手 2 人を比較しどちらが上位になるのかをニューラルネットワークに学習させ、ニューラルネットワークの出力値を各選手について集計することによって、車券的中確率を計算した。有効であると考えられる特徴量を幾つか追加することによって、予測性能の向上が確認された。車番二連勝単式のオッズを市場の予想勝率に変換し、ニューラルネットワークの入力に使用することで、オッズの入力特徴量としての有効性を確認した。予想記事の自動生成は、テンプレートに修飾文を当てはめるとい

手法によって、ラインに関する記述、選手個人に注目した記述、車券予想の記述を含む予想記事を生成した。

今後の課題としては、有効な入力特徴量のさらなる探索によって予測精度でオッズを超えることを目指す。また、最終オッズを予測に使用した場合の有効性を確認したが、実レースの予測を行う際には最終オッズを使用することはできない。投票締め切り前のオッズを使用した場合の性能変化を調査し、締め切り前のオッズの有効性を調査する必要がある。記事生成については、文テンプレートに当てはめる修飾文の設定や、修飾文選択の条件の設定による記事に出現する語彙の増加が今後の課題である。また、現時点では対象外としたレース展開についての記述も今後の課題であり、過去の決まり手を用いた予測や、マルチエージェントシミュレーション等のアプローチによって実現可能であると考えている。

謝辞

本研究は、株式会社チャリ・ロトの支援を受け実施されたものです。ここに感謝の意を表します。

参考文献

- [1] 村上聡一郎, 笹野遼平, 高村大也, 奥村学: 打者成績からのインニング速報の自動生成, 言語処理学会第 22 回年次大会発表論文集, 2016.
- [2] 田川裕輝, 嶋田和孝: テンプレートの自動生成によるインニングの要約文生成, 言語処理学会第 23 回年次大会発表論文集, 2017.
- [3] 村上聡一郎, 渡邊亮彦, 宮澤彬, 五島圭一, 柳瀬利彦, 高村大也, 宮尾祐介: 時系列数値データからの概況テキストの自動生成, 言語処理学会第 23 回年次大会発表論文集, 2017.
- [4] 村上聡一郎, 笹野遼平, 高村大也, 奥村学: 数値予報マップからの天気予報コメントの自動生成, 言語処理学会第 23 回年次大会発表論文集, 2017.
- [5] Davoodi, E. and Khanteymooi, A. R.: Horse Racing Prediction Using Artificial Neural Networks, *Proceedings of the 11th WSEAS International Conference on Neural Networks and 11th WSEAS International Conference on Evolutionary Computing and 11th WSEAS International Conference on Fuzzy Systems*, NN'10/EC'10/FS'10, Stevens Point, Wisconsin, USA, World Scientific and Engineering Academy and Society (WSEAS), pp. 155-160 (online), available from <http://dl.acm.org/citation.cfm?id=1863431.1863457> (2010).
- [6] Edelman, D.: Adapting support vector machine methods for horserace odds prediction, *Annals of Operations Research*, Vol. 151, No. 1, p. 325 (online), DOI: 10.1007/s10479-006-0131-7 (2006).
- [7] オッズパーク株式会社: 競輪 (KEIRIN・ケイリン) ならオッズパーク競輪 | 予想情報も充実! (オンライン), 入手先 <http://www.oddspar.com/keirin/> (参照 2018-05-29).
- [8] Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, *CoRR*, Vol. abs/1412.6980 (online), available from <http://arxiv.org/abs/1412.6980> (2014).