

競輪における予想記事生成のためのレース結果予測

吉田 拓海^{1,a)} 横山 想一郎² 山下 倫央² 川村 秀憲²

受付日 2019年1月28日, 採録日 2019年7月3日

概要: 長年減少していた競輪の車券売上はインターネット投票の売上増加の影響によって、近年では回復傾向にある。そこで、競輪事業を活性化するためには Web 上での積極的な情報発信が必要である。競輪の情報発信の例として予想記事があげられる。予想記事は専門家によるレース結果の予想が記述された記事であり、情報発信として有効である。現在、競輪の予想記事は全国で毎日実施されている約 60 レースの各レースについて人手で執筆されている。また、前日のレース結果に応じて参加選手が決定されるということもあり、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。本研究の目的は、レース開始前の情報から予想記事を自動生成することである。記事の自動生成のために、機械学習によりレース結果を予測し、予測結果を説明する記事をテンプレートによって生成した。

キーワード: 記事生成システム, 機械学習, 予測

Prediction of Race Results for Generation of Predicted Articles in Keirin

TAKUMI YOSHIDA^{1,a)} SOICHIRO YOKOYAMA² TOMOHISA YAMASHITA² HIDENORI KAWAMURA²

Received: January 28, 2019, Accepted: July 3, 2019

Abstract: Keirin ticket sales which had been decreasing for many years have been recovering in recent years because of increase sales on the Internet. So, proactive information provision is necessary to activate the Japanese Keirin business. There is an article describing prediction of the race result as example of information provision. However, the burden of writing articles by hand is great. This is because the number of races per day is about 60 races and it is very large. Another reason is that participating cyclists will be decided according to the race result of the previous day. Therefore, there is a strong demand for a technique for automatically generating content including articles. The purpose of this research is to automatically generate articles from the information obtained before the start of the race in Japanese Keirin. In order to automatically generate articles, we predict race results by machine learning. And, we generate an article explaining the predicted results using the template.

Keywords: article generation system, machine learning, prediction

1. はじめに

スポーツの分野においては、記事の自動生成に関する研究が積極的に行われている。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、テキスト速報から野球ダイジェストを生成する研究 [2]、打者成績からインニングの要約文を生成する研究 [3] などが行われている。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生成する研究 [4] や、天気予報コメントを自動生成する研究 [5] が行われている。上記の研究は、人手によるテキスト生成の負担を軽減するために行われている。

競輪においても文章自動生成技術の高いニーズがある。2000 年以前から減少し続けていた競輪の車券売上がイン

¹ 北海道大学大学院情報科学研究科
Graduate School of Information Science and Technology,
Hokkaido University, Sapporo, Hokkaido 060-0814, Japan

² 北海道大学大学院情報科学研究科
Faculty of Information Science and Technology, Hokkaido
University, Sapporo, Hokkaido 060-0814, Japan

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

ターネット投票の売上増加の影響によって回復傾向にあり [6], 競輪事業活性化のために Web 上での情報発信が強く求められている。競輪の情報発信の例として専門家によるレース結果の予想が記述された予想記事があげられる。競輪の1日あたりのレース数は約60レースであり, その各レースについて現在は予想記事が人手で執筆されている。また, 前日のレース結果に応じて参加選手が決定されるということもあり, 記事の生成の負担が大きい。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。

このような背景をふまえて, 本研究の目的を競輪の予想記事をレース開始前の情報から自動生成することとする。文献 [1], [2], [3], [4] では, 生成の対象となる記事の内容は過去の結果であり, 記事を生成するためのデータがすでに用意されている。文献 [1], [3] では打者成績, 文献 [2] ではテキスト速報, 文献 [4] では株価の時系列データがすでに用意されている。また, 天気予報コメント生成の研究 [5] も生成するコメントの内容は未来の結果だが, 天気予報マップというデータからコメント生成をしており, コメント生成のためのデータがすでに用意されている。一方で本研究では, 生成の対象となる記事の内容は未来の結果であるため, 記事を生成するために必要なデータであるレース結果は用意されていない。

そこで, 本研究では, 以下のように競輪のレース結果の予測結果を用いた記事生成システムを構築する。最初に機械学習を用いて記事生成の対象となる競輪レースの結果をレース開始前の情報から予測する。レース結果として1着から3着の選手を予測する。次に, レース結果予測と選手のデータに基づいて, 事前に用意したテンプレート群から使用する記事テンプレートを選択し, レース結果に基づいてテンプレートに選手の情報を当てはめて, 予想記事を生成する。本稿では, 本システムにおいて重要であるレース結果予測の精度向上のために, 競輪のレース結果予測において有効なモデルについて検討する。また, その予測結果に基づいて予想記事を生成するシステムと生成された記事を示すことで, 本研究の目的が達成されたことを示す。

本稿の貢献は次の2つである。

- (1) 競輪のレース結果予測において有効な予測モデルについて検討し, 有効性の高いモデルを提案した。先行研究である競馬のレース結果予測で一般的に用いられる予測モデルとは異なるモデルを提案し, 提案モデルの有効性を示した。
- (2) レース開始前に取得可能な情報をもとに予想記事を生成するシステムを構築した。機械学習によってレース結果を予測し, 予測結果をもとにテンプレートにより予想記事を生成することによって実用レベルの記事生成システムを構築し, 実社会の課題解決に貢献した。本稿では, 2章で本研究で取り扱う競輪について示す。

3章でレース結果予測のモデルについて示し, 4章で各モデルを比較する実験について示す。5章で記事生成手法について示し, 6章でその結果と考察を示す。最後に7章でまとめと今後の展望について示す。

2. 競輪

本章では, 本研究で取り扱う競輪について, 競技としての競輪の概要と競輪における予想記事について説明する。また, 記事生成システムがどのように予想記事を生成するのかを説明する。

2.1 競技概要

競輪は, 自転車を使用した日本発祥のトラックレースである。レースは基本的に9人で行い, バンクと呼ばれる競争路を周回し1,500mから3,000mの距離を競う。

競輪にはその他の競技にはないラインという要素がある。選手はレース中にラインと呼ばれる縦列を形成し, 同じラインの選手同士で協力しながらレースを進める。出場選手は誰の後ろにつくのか, 自分が先頭で走るか否かについてレース前に宣言するため, どの選手同士がラインを組むかはレース前に公開される情報である。多くの場合は, 同じ地域に所属する選手同士でラインを組む。

ラインの先頭を走る選手(先行選手)は, 走る速度や加速のタイミングを自由に決定することができるが, 風の抵抗をラインの中で最も受けるため, 他の選手よりも体力を消耗する。先行選手の後ろを走る選手(番手選手)は, 前方の選手を風よけに使用することができるため体力を温存することができる。その代わりに, 番手選手は後ろから迫ってくる他ラインの選手をブロックすることで, 先行選手を援護する役割がある。最後のゴール前の直線まで選手は同一ラインの選手と協力して走るが, レースの賞金は個人に与えられるため, 最後は同一ラインの選手とも競争し, 各個人が1位を目指す。ラインによって作られるレース展開は競輪特有のものである。

また, 選手をエージェントとすると, 競輪は協調タスクと競争タスクが混在する複雑なマルチエージェント環境と考えることができる。各エージェントは同一ラインのエージェント同士で協調し, 他のラインと競争する。しかし, 各エージェントの目的は自身の利得の最大化であるため, 協調する同一ライン内のエージェントとも競争する。このように競輪はマルチエージェント環境としても興味深く, 研究の意義がある。

2.2 競輪における予想記事

本節では, 現在人手によって作成されている予想記事にどのような情報が含まれているかを検証する。さらに, どのように予想記事を自動生成するかを説明する。

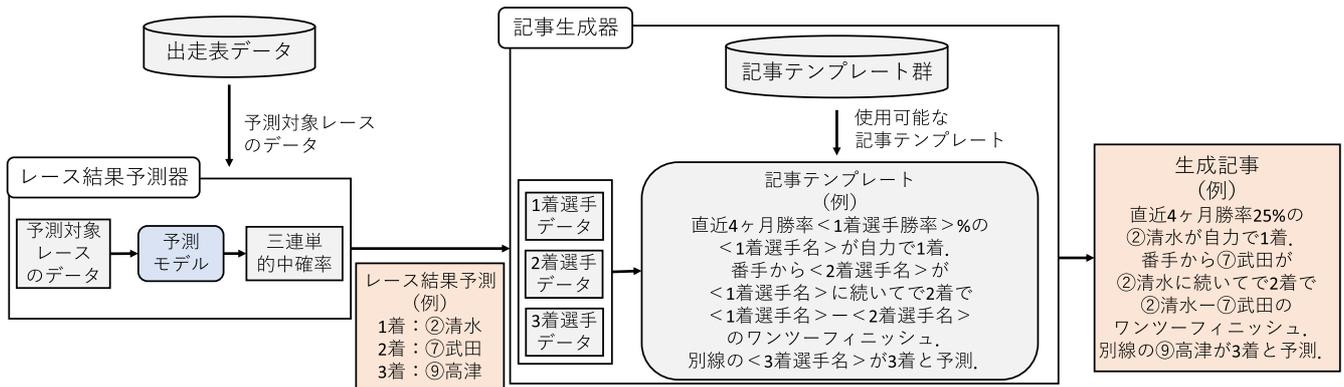


図 1 記事生成システムの概要
Fig. 1 Overview of article generation system.

2.2.1 既存記事

競輪では実施される全レースに対して人手によって予想記事が作成され、新聞や投票サイトでユーザーに向けて配信されている。投票サイトで配信されている既存記事の例を以下に示す。既存記事はオッズパーク [7] より引用した。

- 既存記事例 1
 - 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。
- 既存記事例 2
 - 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
- 既存記事例 3
 - 藤縄-中澤の近畿勢に期待。隅田-高津、中部勢の抵抗も見ものだろう。
- 既存記事例 4
 - 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。

以上から、既存記事が含む情報として次の4つがあげられる。

- 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

上記4つの情報が具体的にどのような情報なのかを、既存記事例1を例に説明する。上位予想選手は、記事の筆者が上位に入る可能性が高いと予想する選手であり、既存記事例1では「野木」、「丸山」、「大久保」、「吉田」である。上位予想選手に関する情報は、なぜその選手が上位に入る

と予想されたかを説明する情報であり、既存記事例1では「鋭さ光る」、「攻め多彩な」である。ラインに関する情報は、ラインによって生じるレース結果についての情報であり、既存記事例1では「磯島が駆けて番手の」、「目標の高鍋次第」である。レース展開に関する情報は、レースの途中経過を表し、既存記事例1では「逆転の捲りに一考」である。

既存記事例3は上位予想選手に関する情報とレース展開に関する情報を含んでいない。このように、必ずしも4つすべての情報が既存記事に含まれているというわけではない。

現時点では、過去のレースのレース中の各選手の順位などのレース展開に関するデータがない。また、既存記事例3のようにレース展開に関する記述がない既存記事も存在するということから、本稿では上記4つの情報のうち、レース展開に関する情報を除いた上位予想選手、上位予想選手に関する情報、ラインに関する情報を含む記事を生成する記事生成システムを構築することを目的とする。

2.2.2 記事生成アプローチ

記事を生成する手法として、テンプレート型生成手法 [8] と Encoder-Decoder モデル [9] による手法の2つがあげられる。

テンプレート型生成手法は古くから用いられている手法であり、文法的に正しい文を生成できる保証があるという利点がある。このテンプレートを自動獲得するという研究 [1], [3] も行われている。この手法は事前にテンプレートとその利用条件を用意するコストを削減することができるが、文法的に正確ではなく、事実とは異なる文を生成してしまう可能性があるという報告 [3] もある。

Encoder-Decoder モデルは、機械翻訳 [10] や画像キャプション生成 [11] などのタスクで研究成果が報告されており、スポーツ記事の生成について適用した研究 [3] もある。ただし、テンプレート自動獲得と同様に、文法的に正確でない、事実とは異なる文を生成してしまう可能性がある。

本稿で提案する記事生成システムには、事前に人手でテ

ンプレートとその利用条件を作成するテンプレート型生成手法を採用する。記事生成システムの概要を図 1 に示す。

本システムは、まずレース結果予測器が、予測対象レースのデータを入力とする予測モデルを用いて、レース結果を予測する。その後、記事生成器が、レース結果予測器の出力したレース結果予測と選手データ、ラインデータに基づいて、事前に作成した記事テンプレート群から使用する記事テンプレートを選択する。選択した記事テンプレートに選手名などの情報を当てはめることで記事を生成する。

3. レース結果予測手法

本章では、機械学習を用いてレース結果を予測する手法について説明する。

予測するレース結果は、1着から3着となる選手の順序である。そのために、車番 i が1着、車番 j が2着、車番 k が3着になる確率 p (1st = i , 2nd = j , 3rd = k) を推定する。競輪において1着、2着、3着を着順通りに予想する賭式を車番三連勝単式 (三連単) と呼ぶ。以降では、確率 p (1st = i , 2nd = j , 3rd = k) を三連単的中確率と呼ぶ。推定された三連単的中確率が最大のものをレース結果予測として、記事生成に使用する。

競輪とよく似た競技として競馬があげられ、競馬に関してはレース結果を予測する研究が数多く行われている。ニューラルネットワークを用いた予測 [12] や、サポートベクターマシンを用いた予測 [13]、ロジスティック回帰を用いた予測 [14]、ファジィ論理を用いた予測 [15] などが行われている。これらの研究では、ある1頭の馬のデータを入力として、その馬が1着になるか否かの分類や走破タイム、マージン (1着の馬との差) の回帰を行い、レース結果を予測している。

本研究で取り扱う競輪と競馬との違いは、馬か自転車か、出走人数、開催頻度など数多くあるが、最大の違いはラインである。ラインによる選手間の相互作用を考慮することは競輪の予測において重要である。

そこで本稿では、先行研究のように選手1人を入力とする1人モデルを提案するだけでなく、選手2人、9人を入力とする2人モデルと9人モデルも提案し、1人モデル、2人モデル、9人モデルの性能を比較する。本稿では、各モデルの分類器の学習アルゴリズムとして、著者らのこれまでの研究 [16] において良い精度を示した多層パーセプトロン (MLP) を用いる。

3.1 1人モデル

1人モデルは先行研究と同様に選手1人を入力として、その選手の強さを推定し、その値からレース結果を予測するモデルである。1人モデルは次の2つのステップによって三連単的中確率を推定する。

(1) 選手1人 (車番 i) を二値分類器の入力とし、車番 i が

1着になる確率を出走選手9人に対して求める。

(2) それらの値から三連単的中確率を近似的に計算する。以降では、二値分類器の入出力と三連単的中確率の計算方法について説明する。

3.1.1 二値分類器の入力/出力

1人モデルはレースの特徴と選手1人 (車番 i) の特徴を入力とする。入力 x_i の出力 y_i を車番 i が1着になる確率として扱う。ここで、 y_i は総和が1になる ($\sum_i y_i = 1$) 保証はないため、式 (1) のような補正を行う。

$$y_i(\text{correction}) = \frac{y_i}{\sum_j y_j} \quad (1)$$

3.1.2 二値分類器の学習

入力された選手が1着になるかを二値分類器に学習させる。入力 x_i の教師ラベル t_i を次のように定義する。

$$t_i = \begin{cases} 0 & \text{place}_i > 1 \\ 1 & \text{place}_i = 1 \end{cases} \quad (2)$$

ここで、 place_i は車番 i の順位である。

3.1.3 三連単的中確率の計算

二値分類器の出力値を用いて三連単的中確率を近似的に計算する。車番 i が1着で車番 j が2着となる条件付き確率を、車番 i がいないレースでの車番 j の1着確率と考える計算手法 [17] により計算を行う。計算式を式 (3) に示す。

$$p(\text{1st} = i, \text{2nd} = j, \text{3rd} = k) = \frac{y_i y_j y_k}{(1 - y_i)(1 - y_i - y_j)} \quad (3)$$

3.2 2人モデル

2人モデルは選手2人を比較することによって、レース結果を予測するモデルである。このモデルは次の2つのステップによって三連単的中確率を推定する。

(1) 選手2人 (車番 i , 車番 j) を二値分類器の入力とし、車番 i が車番 j よりも上位になる確率を出走選手9人の全ペアに対して求める。

(2) それらの値から三連単的中確率を近似的に計算する。以降では、二値分類器の入出力と三連単的中確率の計算方法について説明する。

3.2.1 二値分類器の入力/出力

2人モデルはレースの特徴と選手2人 (車番 i , 車番 j) の特徴を入力とする。入力 $x_{i,j}$ の出力 $y_{i,j}$ は、車番 i が車番 j より上位になる確率として扱う。ここで、出力値は $y_{i,j} \neq 1 - y_{j,i}$ となっているため、集計時に式 (4) のような補正を行う。

$$y_{i,j}(\text{correction}) = \frac{y_{i,j} + (1 - y_{j,i})}{2} \quad (4)$$

3.2.2 二値分類器の学習

入力された2人の選手のどちらが上位になるかを二値分

類器に学習させる。入力 $x_{i,j}$ の教師ラベル $t_{i,j}$ を次のように定義する。

$$t_{i,j} = \begin{cases} 0 & place_i > place_j \\ 1 & place_i < place_j \end{cases} \quad (5)$$

また、着順の上位3人を予測するための学習データとして、4着から9着の選手同士を比較したデータは学習の妨げになることが考えられる。そのため本稿では、4着から9着同士を比較したデータ以外のデータを学習に使用する。後述する比較では、全データを学習に使用したモデルを2人、4着から9着の選手同士を比較したデータを学習に使用しないモデルを2人、学習データ選別と表記する。

3.2.3 三連単的中確率の計算

二値分類器の出力値を用いて三連単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、その積によって三連単的中確率を計算する。

- $p(1st = i)$
 - 車番 i が1着になる確率
- $p(2nd = j | 1st = i)$
 - 車番 i が1着のとき、車番 j が2着になる確率
- $p(3rd = k | 1st = i, 2nd = j)$
 - 車番 i が1着、車番 j が2着のとき、車番 k が3着になる確率

車番 i が1着になる確率は、車番 i がその他の選手全員に勝つ確率、車番 i が1着のとき、車番 j が2着になる確率は、車番 j が車番 i 以外の全員に勝つ確率であるという考えから各確率を式(6), (7), (8)で計算する。また、計算された値の総和が1になる ($\sum_i p(1st = i) = 1$) ように正規化を行う。

$$p(1st = i) = \frac{\prod_{m \neq i} y_{i,m}}{\sum_l (\prod_{m \neq l} y_{l,m})} \quad (6)$$

$$p(2nd = j | 1st = i) = \frac{\prod_{m \neq i,j} y_{j,m}}{\sum_l (\prod_{m \neq i,l} y_{l,m})} \quad (7)$$

$$p(3rd = k | 1st = i, 2nd = j) = \frac{\prod_{m \neq i,j,k} y_{k,m}}{\sum_l (\prod_{m \neq i,j,l} y_{l,m})} \quad (8)$$

3.3 9人モデル

9人モデルは、選手9人を比較することによってレース結果を予測するモデルである。このモデルは、1人モデルや2人モデルのような2段階の処理は行わずに多クラス分類器によって三連単的中確率を直接出力する。

3.3.1 多クラス分類器の入力/出力

9人モデルはレースの特徴と選手9人(車番1, 車番2, ..., 車番9)の特徴を入力とする。対応する選手のない入力値は0とする。このモデルは、多クラス分類器によって三連単的中確率を直接出力する。出力 y は $y[index(i, j, k)] = p(1st = i, 2nd = j, 3rd = k)$ となる504

次元のベクトルである。 $index(i, j, k)$ を式(9)で定義する。

$$index(i, j, k) = (i - 1) \times 56 + (j - 1) \times 7 + (k - 1) - d_1 - d_2 - d_3$$

$$d_1 = \begin{cases} 0 & i > j \\ 7 & i < j \end{cases}$$

$$d_2 = \begin{cases} 0 & i > k \\ 1 & i < k \end{cases}$$

$$d_3 = \begin{cases} 0 & j > k \\ 1 & j < k \end{cases} \quad (9)$$

式(9)は、選手9人から3人の車番を選んで得られる順列 (i, j, k) を3桁の数値 $(100i + 10j + k)$ として昇順に並べることで0から503の値を割り当てている。

3.3.2 多クラス分類器の学習

多クラス分類器に上位3人の選手の順序を学習させる。学習に用いる教師ラベル t は正解の三連単を表すワンホットベクトルであり、 $t[index(1st, 2nd, 3rd)] = 1$ となる。

また、9人モデルの学習データは通常1レースに対して1つであり、1人モデルや2人モデルと比較すると入力の複雑さに対して学習データ数が少ないということが考えられる。そこで本稿では、選手の車番をランダムに入れ替えたデータを学習データに加えることでデータを増量する。本稿ではこれによって学習データ数を8倍に増量した。後述する比較では、データ増量を行わないモデルを9人と表記し、データ増量を行ったモデルを9人_学習データ増量と表記する。

4. 予測手法の比較実験

本章では、前章で提案した各予測手法に実際のレース結果を適用して予測を行う方法を説明するとともに、予測モデルの性能を比較した実験結果を示す。

4.1 実験設定

本節では、学習とテストに使用したデータ、予測に使用する特徴量、予測手法の性能を比較する項目について示す。

4.1.1 データ

2014年から2016年の間に実施されたレースのデータを使用した。2014年と2015年のデータを訓練データとして使用し、2016年のデータはテストデータとして使用した。各レース数は訓練データが40,230レースで、テストデータが15,731レースである。また、訓練データの2割を検証データとして各モデルのハイパーパラメータを決定した。本稿では、以下の条件を満たすデータを使用した。

- 男性レース
- 競りなし

表 1 入力特徴量
Table 1 Input features.

レース特徴	日付 バンクの長さ 同じレースに出走する選手の各個人特徴の平均値 (1人モデル)
個人特徴	年齢 競走得点 1着回数 2着回数 3着回数 1着率 2連対率 バック取り回数 決まり手回数
ライン特徴	先行選手か否か 自身の所属するラインの人数 ライン内で自身の前にいる選手の数 先行選手の個人特徴 (1人, 2人モデル) ラインが同じか否か (2人モデル) ラインベクトル (9人モデル) 前方選手ベクトル (9人モデル) 後方選手ベクトル (9人モデル)

- 同着なし
- 欠損値なし

女性レース（ガールズ競輪）も存在するが、ガールズ競輪にはラインが存在せず競技として男性レースと大きく異なり、レース数も比較的少ないため本稿では対象外とした。競りとは、ライン内で自分が有利な位置につくためにライン内の位置を複数人で奪い合うことを意味する。競りに勝った選手はその位置につくことができるが、負けた選手はその位置につくことができない。そのため、競りがあるレースはラインがレース前に確定しないので、本稿では対象外とした。同着レースのデータは同着なしのデータと同様に教師ラベルを付与することができない。また、同着レースのデータ数も少ないため本稿では対象外とした。欠損値とは、特徴量として過去4カ月の成績などを用いているため、過去4カ月レースに参加していない選手は、その成績がないため欠損値となる。同着レース同様に対象となるデータ数が少なかったため本稿では対象外とした。

4.1.2 特徴量

レースに関するデータ（レース特徴）と選手に関するデータ（選手特徴）を予測に用いた。また、選手に関するデータとして選手個人に関するデータ（個人特徴）と選手の所属するラインに関するデータ（ライン特徴）を用いた。本稿では、インターネット上で誰でも閲覧可能なデータを予測に用いた。特徴量の詳細を表1に示す。

レース特徴はレースに関する特徴であり、日付とバンクの長さを使用した。バンクの長さは333m, 335m, 400m,

500mの4種類がある。1人モデルでは、同じレースに出走する他の選手と比較するための特徴として、同じレースに出走する選手の各個人特徴の平均値を用いた。

個人特徴は、選手個人についての特徴である。年齢以外の個人特徴は選手の強さを推定するのに有効であると考えられる特徴を採用しており、過去4カ月の成績データに基づいて算出されている。競争得点はレースの成績によって選手に与えられる点である。バック取り回数は最終半周の時点で先頭を走っていた回数のことである。決まり手はレースの決着のつき方であり、逃げ・捲り・差し・マークの4種類があり、決まり手回数はそれぞれの回数のことである。

ライン特徴は、ラインの影響を考慮するための特徴量である。ライン特徴に関しては、1人モデル、2人モデル、9人モデルのそれぞれで異なる特徴量が用いられる（一部同じ特徴量を用いる）。先行選手か否か、自身の所属するラインの人数、ライン内で自身の前にいる選手の数に関しては、3つの予測モデルで同じ形式の特徴量が利用される。先行選手か否かは、先行選手の場合は1、そうでない場合は0となる。

先行選手の個人特徴は、1人モデルでは先行選手の個人特徴を用いた。2人モデルでは先行選手の個人特徴と比較する2人の選手が同じラインか否かを表す特徴を用いた。

ラインが同じか否かは、2人モデルのみに用いられて、選択された2人が同じラインの場合は1、それ以外は0となる。ラインベクトル、前方選手ベクトル、後方選手ベクトルは9人モデルのみに用いられ、自身のラインと前後の選手を表すためのベクトルで表現される。たとえば、ラインが1-9-3の場合、車番9のラインベクトルは $[1, 0, 1, 0, 0, 0, 0, 0, 1]$ 、前方選手ベクトルは $[1, 0, 0, 0, 0, 0, 0, 0, 0]$ 、後方選手ベクトルは $[0, 0, 1, 0, 0, 0, 0, 0, 0]$ となる。前方または後方の選手がいない場合はゼロベクトルを用いた。

後述する比較では、各モデルで共通しているライン特徴のみを使用したモデルを比較に使用し、1人共通ライン特徴のように表記する。9人モデルについては1人、2人モデルと同じ先行選手の個人特徴をライン特徴として使用したモデルも比較に使用し、9人先行ライン特徴と表記する。

各特徴量は最大値1、最小値0に正規化して入力に使用した。1人モデルはレース特徴と1人の選手特徴を結合した42次元ベクトル、2人モデルはレース特徴と2人の選手特徴を結合した62次元ベクトル、9人モデルはレース特徴と9人の選手特徴を結合した381次元ベクトルを入力ベクトルとした。

4.1.3 比較

各種モデルについて、1着選手、二車単、三連単の的中精度、二車単と三連単の回収率、三連単の的中確率の推定誤差を比較した。二車単とは1着、2着を着順通りに予想する賭式である。的中精度については、Top-K-accuracyを

表 2 予測モデルの比較 (1 着, 二車単, 三連単の Top K accuracy)
 Table 2 Comparison of prediction model (top K accuracy for the 1st, 2nd, 3rd cyclists).

	1 着 @1 [%]	1 着 @2 [%]	1 着 @3 [%]	二車単 @1 [%]	二車単 @5 [%]	二車単 @10 [%]	三連単 @1 [%]	三連単 @5 [%]	三連単 @10 [%]
1 人	41.8	62.8	76.2	13.7	42.7	60.0	3.3	12.7	20.9
1 人_共通ライン特徴	41.2	62.6	76.1	13.6	42.4	60.0	3.3	12.6	20.7
2 人	39.4	60.4	73.2	15.7	43.4	59.5	4.8	17.1	26.7
2 人_共通ライン特徴	39.2	59.9	73.2	14.3	41.9	58.8	4.1	15.5	24.7
2 人_学習データ選別	41.7	62.6	75.6	16.7	45.5	61.6	5.1	17.6	27.4
9 人	36.8	57.7	72.4	17.9	46.0	62.3	6.3	18.8	28.4
9 人_共通ライン特徴	38.3	60.0	73.6	13.7	41.7	58.4	3.4	13.6	22.9
9 人_先行ライン特徴	36.8	57.1	70.9	7.8	26.5	42.9	2.7	11.5	29.3
9 人_学習データ増量	40.6	61.8	75.5	19.1	50.4	66.8	7.7	22.8	33.4
オッズ	—	—	—	21.2	52.6	68.9	9.1	26.2	37.2

用いた。1 着選手については $K = 1, 2, 3$, 二車単と三連単については $K = 1, 5, 10$ の Top-K-accuracy を用いた。回収率は的中確率が最大の車券を各レースで 1 点ずつ購入した場合の総配当金を総購入費で除算した値であり、配当の計算には確定オッズを使用した。三連単の的中確率の推定誤差は次の手順で計算した。

- (1) 三連単の的中確率の推定値の小数点第 3 位以下を切り捨てる。
- (2) 小数点第 3 位以下を切り捨てた各数値ごと (0.00, 0.01, 0.02, ..., 0.99) に、正解の三連単をもとに実際の三連単の的中確率を計算する。
- (3) その平均絶対誤差 (MAE) を計算する。

三連単の的中確率は値が高くなるにつれて発生頻度も少なくなるため、本稿では 0.2 以下の三連単の的中確率を誤差計算の対象とした。また、横軸を推定確率、縦軸を実際の的中確率としたグラフによりその関係を示す。

予測のベースラインとして、確定オッズによる予測を用いた。オッズとは公営競技における払戻金の倍率であり、投票券の投票数によって決定される。投票数の多い (人気の高い) 投票券のオッズは低く、投票数の少ない (人気の低い) 投票券のオッズは高くなる。つまり、オッズは投票者全体の予想を反映していると考えられる。

4.2 結果・考察

各モデルとオッズの 1 着予測, 二車単予測, 三連単予測の Top-K-accuracy を表 2 に示す。各モデルの計算する三連単の的中確率の推定の平均絶対誤差を表 3 に示し、その中の 1 人, 2 人, 9 人について、推定値と真値についての関係を図 2 に示す。各モデルとオッズの二車単, 三連単の回収率を表 4 に示す。

4.2.1 ライン特徴についての比較

各モデルについて共通のライン特徴のみを使用するモデルとモデル特有のライン特徴を使用するモデルを比較する。表 2 から 1 人モデルについては、先行選手の個人特

表 3 予測モデルの比較 (三連単の的中確率の推定誤差)

Table 3 Comparison of prediction model (MAE of estimated probability).

	MAE
1 人	0.037
1 人_共通ライン特徴	0.038
2 人	0.054
2 人_共通ライン特徴	0.059
2 人_学習データ選別	0.061
9 人	0.010
9 人_共通ライン特徴	0.051
9 人_先行ライン特徴	0.046
9 人_学習データ増量	0.018

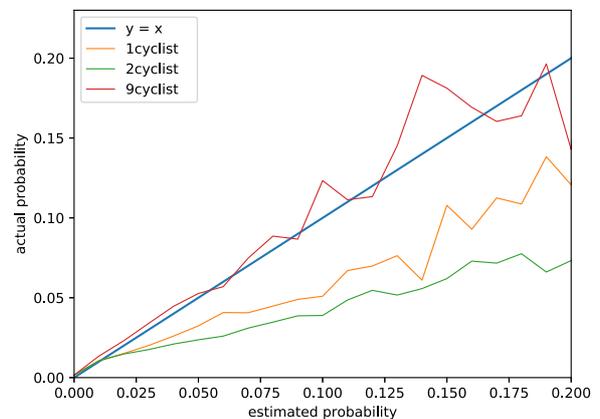


図 2 予測モデルの比較 (三連単の的中確率の推定精度)

Fig. 2 Comparison of prediction model (accuracy of estimated probability).

徴をライン特徴に使用することによる精度の違いは確認できない。2 人モデルについては、先行選手の個人特徴をライン特徴に使用することで、1 着については精度の違いは確認できないが、二車単, 三連単の精度の向上が確認できる。二車単, 三連単というラインの影響を強く受ける項目における精度の向上は、使用したライン特徴が競輪固有の問題であるラインの扱いにおいて有効であることを示して

表 4 予測モデルの比較 (回収率)

Table 4 Comparison of prediction model (recovery rate).

	二車単回収率 [%]	三連単回収率 [%]
1 人	83.6	74.0
1 人_共通ライン特徴	83.3	71.6
2 人	82.1	71.8
2 人_共通ライン特徴	82.0	81.8
2 人_学習データ選別	79.3	69.6
9 人	89.0	92.1
9 人_共通ライン特徴	79.5	71.3
9 人_先行ライン特徴	75.0	73.9
9 人_学習データ増量	83.9	93.2
オッズ	78.4	81.5

いる。9人モデルについては、ラインを表現する3種のベクトルをライン特徴に使用することで二車単、三連単の精度が向上しており、使用した特徴のラインの扱いにおける有効性を示しているが、1着については精度が低下している。1着予測という二車単、三連単と比較してラインの影響の小さい単純なタスクにおいては、ライン特徴追加による入力の複雑化のデメリットが大きかったと考えられる。9人モデルにおいて、1人モデルらと同様に先行選手の個人特徴をライン特徴に使用すると、精度が著しく低下することからも、9人モデルの二車単、三連単の精度について、ライン特徴として使用したベクトルの貢献が大きいことが分かる。

4.2.2 各モデルについての比較

表2から各モデル(1人, 2人, 9人)を比較すると、二車単、三連単については、9人モデルが最も精度が高いことが確認できる。三連単については表3, 図2から、9人モデルが最も正しく確率を推定可能であることが確認できる。一方で、共通のライン特徴を使用したモデルを比較すると、9人モデルの精度は低いことが確認できる。

9人モデルは出走選手全員のデータを1つの入力として扱うことで、その選手らの関係性を表現するベクトルを入力に取り入れることが可能である。そのため、ラインを考慮した予測が他モデルより高精度で行えたものと考えられる。その一方で、9人モデルの1着精度がライン特徴の追加によって低下したことや、1着の精度については1人モデルが最も精度が高いことから、9人モデルの出走選手全員のデータを1つの入力とする方法には改善の余地があると考えられる。具体的には、1人モデルの枠組みを用いた、各選手単位での特徴抽出による選手のベクトル表現に対して、9人モデルの枠組みを適用することで、さらなる精度向上が期待できる。

2人モデルはどの項目においても最高精度を示すことはない。しかし、共通のライン特徴を用いた場合の三連単の精度については、最も精度が高いことが確認でき、選手を1対1で比較するという枠組みを用いて得られる特徴は予

測に有効と考えられる。今後は、選手のベクトル表現を得る際に、2人モデルの枠組みを組み込むことも検討する。

4.2.3 学習データ選別・増量についての比較

2人モデルと9人モデルの学習データ選別、学習データ増量について比較すると、どちらも選別、増量を行わないモデルよりも精度の向上が確認された。特に9人モデルのデータ増量は、1着のTop-K-accuracyを約4%向上させるなど、大きく精度を向上させた。一方で表3の確率の推定誤差に注目すると、データ選別、データ増量を行ったモデルは誤差がわずかに増加している。データ増量は的中精度の面では非常に効果的であったが、確率の推定精度に及ぼす影響について今後調査する必要がある。また、適切なデータ増量数の調査も今後の課題である。

4.2.4 オッズについての比較

オッズと各モデルを比較すると、オッズの予測精度が高いことが確認できる。競馬ではオッズの馬の勝利確率の推定精度が非常に高いという報告[18]もあり、オッズは優れた予測器であるともいえる。しかし、競輪の投票券の購入者は利益を得ることを目的としているため、予想記事生成のための予測には予測精度だけではなく回収率という指標も考慮する必要がある。表2の結果はオッズが最小の投票券的中率であり、その配当は低い。また表4に示したように機械学習による予測(特に9人モデル)は、オッズによる予測よりも高い回収率を示す。

またオッズは、観客の投票がなければ決定されない。出走選手は基本的にはレース前日の夜には公開されており、その時点ではオッズを用いて予測を行うことはできない。

また、オッズは観客の投票によって決定されるため、投票締切前まで変動し続ける。そのため、注目度の低く票のあまり集まらないレースでは、オッズによる予測結果は時間経過にともなって大きく変動することもある。一方で、本稿で提案した予測手法の利点として、本稿で提案した予測手法は時間経過にともなって変動することではなく、投票がなくても予測結果を出力可能であることがあげられる。このように、本稿で提案した予測手法は、的中率に限ればオッズに劣っているものの、回収率、予測結果の安定性、予測結果を出力可能なタイミングといった条件を考慮すると、記事生成のための予測としてオッズより優れていると考えられる。

5. 予想記事の生成手法

本章では、前章で得られたレースの予測結果を用いて予想記事を生成する手法について説明する。2章で示したように、本システムは予測結果、選手データ、ラインデータに基づいて、事前に用意した記事テンプレート群から使用する記事テンプレートを選択し、その記事テンプレートに選手名などの情報を当てはめることで記事を生成する。本稿では、選手名と順位を説明するだけの単純な文をもと

に、記事テンプレートを作成した。以降では、2章で設定した次の3つの情報を含む記事テンプレートを作成する手法と、記事生成時にどのように記事テンプレートを選択するかを説明する。

- 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

5.1 記事テンプレートの作成

次の基本本文をもとに記事テンプレートを作成する。

- <選手名> が <順位> 着.

記事テンプレートは次のように基本本文を3つつなげることで作成する。

- <1着選手名> が1着. <2着選手名> が2着. <3着選手名> が3着.

この文に予測結果を当てはめることで、上位予想選手についての記事を生成可能である。さらに、基本本文の先頭、<順位>の前(中間)、末尾に選手情報のテンプレートとライン情報のテンプレートを挿入することで、上位予想選手に関する情報とラインに関する情報を含む記事テンプレートを作成する。つまり、選手情報のテンプレートとライン情報のテンプレートについて使用条件と挿入可能位置を設定することで記事テンプレートを作成する。

5.1.1 選手情報のテンプレート

選手情報のテンプレートとはどのようなものかを示す。実際に設定した選手情報のテンプレートの例を以下に示す。

- 直近4カ月の勝率<1着選手勝率>%の
 - 挿入可能位置:1着, 先頭
 - 使用条件:1着選手の過去4カ月の勝率 $\geq 20\%$

この選手情報手のテンプレートは、1着選手の基本本文の先頭に挿入可能で、1着選手の過去4カ月の勝率が20%以上の場合に使用可能である。本稿では簡単のため選手情報のテンプレートの挿入可能位置を1着選手の基本本文の先頭に限定した。このように、選手データを使用条件に用い、その条件に応じた出力をするテンプレートを設定することで、上位選手に関する情報を含む記事テンプレートを作成する。

上記の例以外に、本稿では以下の選手データを用いて選手情報のテンプレートを設定した。

- 過去4カ月の1着回数
- 過去4カ月の2連対率
- 年齢

5.1.2 ライン情報のテンプレート

ライン情報のテンプレートとはどのようなものかを示す。実際に設定したライン情報のテンプレートの例を以下に示す。

- でワンツーフィニッシュ
 - 挿入可能位置:2着, 末尾

– 使用条件:1着選手と2着選手が同じライン
このライン情報のテンプレートは、2着選手の基本本文の末尾に挿入可能で、1着選手と2着選手が同じラインの場合に使用可能である。このように、ラインデータを使用条件に用い、その条件に応じた出力をするテンプレートを設定することで、ラインに関する情報を含む記事テンプレートを作成する。

5.2 記事の生成

本節では実際に記事の生成手順について説明する。

記事を生成するときは、記事テンプレート群から予測結果、選手データ、ラインデータに基づいて使用条件を満たしている記事テンプレートを選択する。記事テンプレートの使用条件は、その記事テンプレート内の選手情報のテンプレートとライン情報のテンプレートの使用条件の論理積で表される。つまり、記事テンプレート内のすべての選手情報とライン情報のテンプレートの使用条件を満たしている場合にその記事テンプレートは使用可能となる。そして、選択した記事テンプレートに選手の名前や勝率などのデータを当てはめることで記事を生成する。

6. 記事生成結果・考察

2016年のレースに対し実際に生成した記事の例と既存記事を表5、表6、表7、表8に示す。既存記事はオッズパーク [7] より引用した。

既存記事と比較すると、生成記事は既存記事と同じ選手を上位予想選手としていることが多く確認できる。表5の

表5 生成記事と既存記事の例1

Table 5 Example of generated article and existing article (1).

2016-1-20 佐世保競輪場 第6レース	
既存記事	混戦だが②武藤の自力に期待。③米嶋⑨江口が逆転狙い。⑧砂川も②武藤マークから連入も。穴は⑦高谷の自在戦から。①太田のコース注意。
生成記事	勝ち星を量産している②武藤が自力で1着。別線から①太田が2着とスジ違いの決着。九州ラインから⑨江口が3着と、各ラインから上位者が出る予予測。

表6 生成記事と既存記事の例2

Table 6 Example of generated article and existing article (2).

2016-5-27 佐世保競輪場 第6レース	
既存記事	捲りタイプの⑤越智だが前々から自力で首位期待。②蓮井の逆転も。⑦佐伯も同県③渡邊マークから7=3狙いは十分。ここは押さえない。
生成記事	直近4カ月勝率22.7%の②蓮井が⑤越智の自力から差して1着。同ラインの⑤越智が②蓮井に続いて2着でワンツーフィニッシュ。④桜井が3着。

表 7 生成記事と既存記事の例 3

Table 7 Example of generated article and existing article (3).

2016-10-22 青森競輪場 第3レース	
既存記事	△村本に期待 村本が自力含みで臨機応変に攻める。小林は村本に付けて好位差し。南関のワンツー本命。武笠も狙いたい一車。小塚が駆ければ首位浮上。好配は線違い。
生成記事	ここ4カ月勝率44.8%の④村本が自力を生かして1着。同ラインの①小林が④村本に続いて2着。⑦清水が3着。南関ラインが別線を抑えて上位独占と予測。

表 8 生成記事と既存記事の例 4

Table 8 Example of generated article and existing article (4).

2016-7-13 岐阜競輪場 第1レース	
既存記事	互角の争いも①伊藤に期待。好機に仕掛けて押し切ろう。南関勢⑨松江が喰い下がる。新鋭⑥竹元が果敢に主導権を握ると、③川島のガードで逃げ切りも十分。
生成記事	若手の⑥竹元が自力を生かして1着。番手から③川島が⑥竹元に続いて2着。⑧西島が3着。

例では、生成記事に記載されている武藤、太田、江口の3人が既存記事にも上位予想選手として記載されている。これは、既存記事を作成した専門家に近い予測ができているということを示している。

生成記事例から上位予想選手に関する情報を含む記事を生成できていることが確認できる。表6の例では、1着予測選手(蓮井)の過去4カ月の勝率が22.7%であるという記述がされている。本稿で設定した条件によって、2016年レースの71.7%に上位予想選手に関する情報を含む記事が生成された。

生成記事例からライン構成を説明する記事が生成できていることが確認できる。表6の例は、1着予測の選手(蓮井)と2着予測の選手(越智)が同じラインに所属していることを説明している。同様に表5の例は、1着から3着予測の選手(村本、小林、清水)が全員同じラインに所属していることを説明している。また、表7の例では、1着から3着予測の選手(武藤、太田、江口)が全員異なるラインに所属していることを説明している。

本稿で設定したライン情報のテンプレートの組合せによる記事テンプレートの総パターン数は438である。選手情報のテンプレートとの組合せも合わせると約4,000になり、競輪の1日のレース数が60ということから十分な数の記事パターン数と考えられる。しかし、もともととなっている基本文が1つということや、同じような単語・フレーズの使い回しにより、既存記事と比較すると語彙や文のパターンが乏しいということが課題である。

生成記事を車券販売サービス会社の関係者に確認してもらった結果、高評価を受け、本システムはWebページ上

で実用化されている。これは、生成記事が実用レベルの質を確保していることを示している。また、本研究の成果は競輪業界の抱える課題の解決に貢献していることを示している。

本稿で提案したレース結果予測を含む記事生成システムは実用化されており、株式会社チャリ・ロトのWebページAI-win^{*1}にて生成記事を投票締切10分前に公開している。また、その過程で得られる各車券の推定的中確率と公開情報であるオッズを用いて、配当金を考慮した購入車券の提案も生成記事と同じく公開している。現在実用中のシステムのレース結果予測には、2人モデルを使用しているが、本稿の結果を受け9人モデルへの移行も検討中である。

7. まとめと今後の展望

本稿では、機械学習によって競輪のレース結果を予測し、予測結果に基づく記事生成システムを提案した。レース結果の予測に関しては、先行研究である競馬予測で用いられている各選手の強さを推定する手法(1人モデル)に加えて、選手2人、9人を直接比較する手法(2人モデル、9人モデル)を提案し、各手法の利点を示した。

また、予測結果に基づいて事前に人手で作成したテンプレートを選択するという競輪の予想記事を自動的に生成するシステムを構築した。テンプレート型生成手法は単純な方法ではあるが、既存記事として必要とされる情報を含み、十分と考えられる記事パターン数を持つことを示した。

今後の課題として、予測に関しては、さらなる精度向上が課題である。現状では確定オッズによる予測の方が高い精度を示している。本稿で使用した特徴量はインターネット上で公開されているデータの一部であり、特徴量探索により精度の向上が期待できる。また、9人モデルが1人モデルより二車単、三連単の予測精度が高いにもかかわらず、1着の予測精度については1人モデルよりも低かったことから、ニューラルネットワークの構造にも改善の余地があると考えられる。これについては、再帰型ニューラルネットワークの適用なども検討しており、現時点でRNNの適用によりMLPによる予測と同等の精度を記録している。ネットワーク構造のさらなる改良によって精度の向上が期待できると考えている。

記事生成に関しては、既存記事と比較すると語彙や記事のパターンが乏しいという課題がある。本稿では、正しい文が出力される保証がある事前に人手でテンプレートを作成するという手法を用いたが、今後はテンプレートの自動獲得やEncoder-Decoderモデルによる生成も視野に入れる必要があると考えている。

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。

*1 <https://ai.chariloto.com/>

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, pp.338-341 (2016).
- [2] 岩永朋樹, 西川 仁, 徳永健伸: テキスト速報を用いた野球ダイジェストの自動生成, 言語処理学会第 22 回年次大会発表論文集, pp.238-241 (2016).
- [3] 田川裕輝, 嶋田和孝: スポーツ要約生成におけるテンプレート型手法とニューラル型手法の提案と比較, 自然言語処理, Vol.25, No.4, pp.357-391 (オンライン), DOI: 10.5715/jnlp.25.357 (2018).
- [4] 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, pp.1117-1120 (2017).
- [5] 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, pp.1121-1124 (2017).
- [6] 経済産業省製造産業局車両室: 競輪・オートレースを巡る最近の状況について (2018), 入手先 (http://www.meti.go.jp/shingikai/sankoshin/seizo_sangyo/sharyo_kyogi/pdf/010.01.00.pdf).
- [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!, 入手先 (<http://www.oddsparc.com/keirin/>) (参照 2018-05-29).
- [8] McKeown, K. and Radev, D.R.: Generating Summaries of Multiple News Articles, *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '95*, pp.74-82, ACM (online), DOI: 10.1145/215206.215334 (1995).
- [9] Sutskever, I., Vinyals, O. and Le, Q.V.: Sequence to Sequence Learning with Neural Networks, *Advances in Neural Information Processing Systems 27*, Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D. and Weinberger, K.Q. (Eds.), Curran Associates, Inc., pp.3104-3112 (2014) (online), available from (<http://papers.nips.cc/paper/5346-sequence-to-sequence-learning-with-neural-networks.pdf>).
- [10] Luong, T., Pham, H. and Manning, C.D.: Effective Approaches to Attention-based Neural Machine Translation, *Proc. 2015 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, pp.1412-1421 (online), DOI: 10.18653/v1/D15-1166 (2015).
- [11] Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., Zemel, R. and Bengio, Y.: Show, Attend and Tell: Neural Image Caption Generation with Visual Attention, *Proc. 32nd International Conference on Machine Learning*, Bach, F. and Blei, D. (Eds.), *Proc. Machine Learning Research*, Vol.37, pp.2048-2057, PMLR (2015) (online), available from (<http://proceedings.mlr.press/v37/xuc15.html>).
- [12] Davoodi, E. and Khanteymoori, A.R.: Horse Racing Prediction Using Artificial Neural Networks, *Proc. 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*, World Scientific and Engineering Academy and Society (WSEAS), pp.155-160 (2010) (online), available from (<http://dl.acm.org/citation.cfm?id=1863431.1863457>).
- [13] 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).
- [14] Silverman, N. and Suchard, M.: Predicting Horse Race Winners Through a Regularized Conditional Logistic Regression with Frailty, *Journal of Prediction Markets*, Vol.7, No.1, pp.43-52 (2013) (online), available from (<https://EconPapers.repec.org/RePEc:buc:jpredm:v:7:y:2013:i:1:p:43-52>).
- [15] Jogeeah, M., Kumar Chandoo, A., Paupiah, S. and Pudaruth, S.: Using Fuzzy Logic to Predict Winners in Horseraces at the Champ de Mars, *Proc. 3rd International Conference on Digital Information Processing, E-Business and Cloud Computing, Reduit, Mauritius 2015* (2015).
- [16] 吉田拓海, 横山想一郎, 山下倫央, 川村秀憲: 競輪予想記事の自動生成に向けた深層学習によるレース結果予測, *SIG-DOCMAS*, Vol.15, No.1, pp.1-7 (2018).
- [17] Harville, D.A.: Assigning Probabilities to the Outcomes of Multi-Entry Competitions, *Journal of the American Statistical Association*, Vol.68, No.342, pp.312-316 (online), DOI: 10.1080/01621459.1973.10482425 (1973).
- [18] 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).



吉田 拓海 (学生会員)

2018年北海道大学工学部情報エレクトロニクス学科卒業。現在、同大学大学院情報科学研究科修士課程在学中。機械学習を用いたスポーツの勝敗予測に関する研究に従事。人工知能学会の会員。



横山 想一郎 (正会員)

2016年3月北海道大学大学院情報科学研究科情報理工学専攻博士後期課程期間短縮修了。博士(情報科学)。同年4月日本学術振興会特別研究員(PD)。2017年2月同大学助教となり現在に至る。スケジューリング、組合せ最適化、機械学習の研究に従事。人工知能学会等の会員。



山下 倫央 (正会員)

2002年北海道大学大学院工学研究科システム情報工学専攻博士後期課程修了。博士(工学)。2003年まで日本学術振興会特別研究員(DC1)。2003年産業技術総合研究所サイバーアシスト研究センター特別研究員。2016年同所人工知能研究センター主任研究員。2017年より北海道大学大学院情報科学研究科准教授。社会システムシミュレーションの開発と社会実装に関する研究に従事。人工知能学会等の会員。



川村 秀憲 (正会員)

2000年北海道大学大学院工学研究科システム情報工学専攻博士後期課程修了。博士(工学)。同年同大学助手。2006年同准教授。2016年同教授となり現在に至る。人工知能, マルチエージェントシステム, 観光情報学等の研究に従事。人工知能学会, 電子情報通信学会, ロボット学会, 観光情報学会等の会員。