

# レーシングゲーム実況テキストモデリングのための 運動力学的素性

石垣 達也<sup>1,a)</sup> 上田 佳祐<sup>1,2,b)</sup> 小林 一郎<sup>1,3,c)</sup> 宮尾 祐介<sup>1,2,d)</sup> 高村 大也<sup>1,e)</sup>

**概要:** 「ターン 2, 曲がっていけるか?」。レーシングゲームなどの映像にはしばしばこのような実況が、音声や字幕として付与され、視聴者の実況対象に対する理解を助ける。本稿では実況発話生成、プランニング、発話分類の 3 つの実況発話モデリング問題に対し、運動力学に基づく素性を新たに提案する。実況において、実況者はサーキット上で起こるイベントを正しく認識し、発話する。従来、レーシングカーの速度やハンドル角度といったセンサーデータから取得した生データを入力として与え、ニューラルネットワークを用い、言語生成問題や分類問題を解く手法が用いられてきた。これらのモデリング問題をより高性能に解くためには、レーシングカーの走行位置がコーナーであるか否か、加減速やレーシングカーの回転といったより深い状況認識が必要となる。しかし、こうした状況解釈をニューラルネットワークに生データから自動で学習させるのは難しい。そこで、本研究では生データに加え、新たに 3 つの素性を用いてモデリングの高度化を目指す。提案素性には外部知識としてサーキットマップを用いる手法、コーナーとストレートでは発話内容を変化させるといった人間の常識知識を運動力学的な計算を用いて表現する手法が含まれる。既存データセットを用いた実験より、言語生成およびプランニングにおいて提案素性の効果を確認した。運動力学に基づく計算により物体の動きを捉え素性とするアイデアは、本稿で扱う実況テキストのモデリング問題のみならず、物体の動きを捉える必要のある天気予報生成など他の言語処理タスクにも汎用的に応用できる可能性がある。

**キーワード:** 言語生成, 発話タイプ同定, プランニング

## Exploiting Kinematic Features for Racing Game Commentary Modeling

TATSUYA ISHIGAKI<sup>1,a)</sup> KEISUKE UEDA<sup>1,2,b)</sup> ICHIRO KOBAYASHI<sup>1,3,c)</sup> YUSUKE MIYAO<sup>1,2,d)</sup>  
HIROYA TAKAMURA<sup>1,e)</sup>

### 1. はじめに

実況は、スポーツ中継やゲーム録画などの映像に音声や字幕として付与され、対象となるスポーツやイベントについて、理解を促す重要な役割を果たす [16]。オンライン上に存在するスポーツやゲームの録画映像以外にも、アマ

チュアスポーツの録画映像などオフラインにも、多くの動画が存在する。しかし、その多くには実況が付与されていない。このような背景から、実況テキストを自動生成する研究 [3], [19], 言語生成の高度化を指向しプランニングを扱う研究 [28] が存在する。また、すでに付与された実況テキストに対する言語理解タスクとして発話分類 [27], [28] が存在し、実況の要約などへの応用を想定している [21], [26]。本稿では実況テキストのモデリング問題として、実況テキスト生成、プランニング、発話分類の 3 つに着目し、運動力学に基づく新たな素性を提案する。

3 つの問題の概要を図 1 に示す。1 つ目の実況発話生成問題では、発話対象時刻が与えられた上でその時刻に発話

<sup>1</sup> 産業技術総合研究所

<sup>2</sup> 東京大学

<sup>3</sup> お茶の水女子大学

a) ishigaki.tatsuya@aist.go.jp

b) ueda.keisuke@aist.go.jp

c) koba@is.ocha.ac.jp

d) yusuke@is.s.u-tokyo.ac.jp

e) takamura.hiroya@aist.go.jp

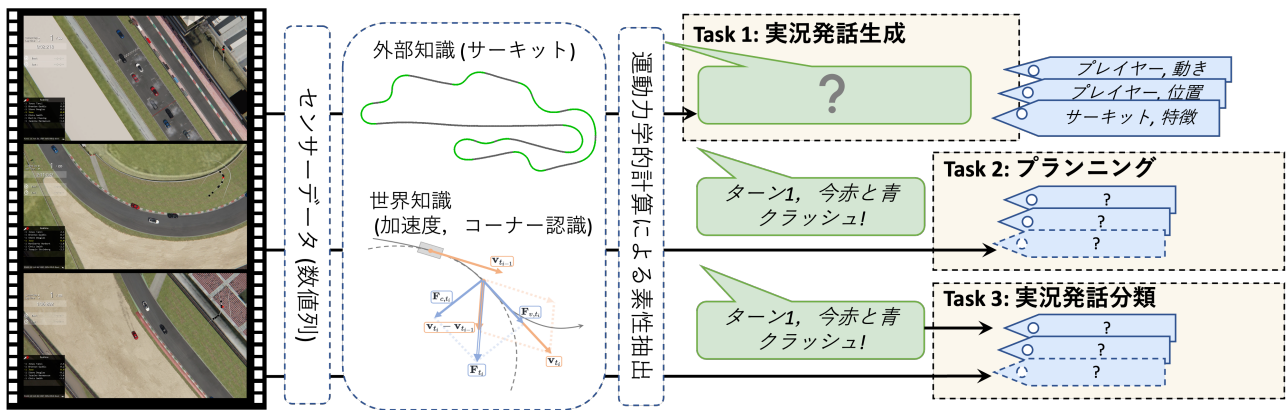


図 1 3つの問題の概要. 生成タスクは時系列数値データを入力とし, テキストを出力する. プランニング問題は生成問題と同じ入力から次々に何を発話すべきかを決定する. 発話分類モデルはこれらの入力の他に発話テキストそのものを受け取り, 発話内容を表現したラベルを付与する

すべき発話テキストを生成する. 実況生成モデルへの入力  
 は, 発話対象時刻よりも過去の時刻での速度やハンドル角  
 度など, センサーから取得した時系列数値データである.  
 2つ目のプランニング問題は実況生成問題と同じ入力デー  
 タを用い, 次に何を発話すべきか表現したラベルを出力す  
 る. ラベルには例えば「サーキットの特徴」「プレイヤーの  
 車のラップタイム」など, 言及対象と言及すべき内容がペ  
 アになった既存の定義ラベル [3] を用いる. 3つ目の発話  
 分類問題は前述した2つの問題と同じ入力に加え, 発話テ  
 キストも入力として受け取り, 同様に定義されたラベルを  
 付与する.

既存研究 [3], [27], [28] では, センサーから取得した時系  
 列数値データをニューラルネットワークによる手法で読み  
 込み, 各下流タスクにおいて用いる. 本研究では, 人間の  
 実況者は生データを用いる際に, より深い状況解釈を加え  
 ていると考える. 例えば, 「プレイヤー, 今ターン2に突  
 入していきます。」という発話であれば, 速度やハンドル  
 角度などの生データを手がかりに, レーシングカーがコー  
 ナーに接近しているか否かを判定するといった状況解釈が  
 行われていると考える. さらに, 人間の実況者は外部知識  
 としてサーキットに関する情報を用い, 「ターン2」という  
 生データには含まれない具体的な位置も捉えている. 実況  
 の言及する内容は実況者の持つ世界知識にも影響される.  
 例えば, コーナーとストレートでは発話内容を変化させたり,  
 レース序盤ではサーキットの特徴についての言及が視  
 聴者から求められており, ゴール周辺ではラップタイムに  
 ついての言及が求められていることを常識知識として知っ  
 ている. このような状況解釈や知識の活用は, 生データか  
 らニューラルネットワークに学習させるのは難しい.

そこで, 本研究ではサーキットマップを外部知識として  
 用い, 運動力学に基づく計算手法によりセンサーから取得  
 した時系列数値データに対しレーシングカーの加減速や

回転といった解釈を明示的に捉える新たな素性を提案す  
 る. 具体的には3つの素性を提案し, それぞれ SECTIONS,  
 CORNERS, ACCELERATION と呼ぶ. 前者2つはサーキット  
 マップを外部知識として用いる手法で, 後者2つは力学的  
 な計算により, コーナーやレーシングカーの加減速を捉え  
 る. SECTIONS はサーキットにおけるレーシングカーのス  
 タートからゴールまでを20分割し, レーシングカーの走行  
 位置をサーキットマップ上の区間に対応づけ素性とする.  
 CORNERS では, 運動力学的な計算として曲率に着目し, サ  
 ーキットマップ上の各点がストレートであるか, コーナーで  
 あるかを自動認識し素性として用いる. ACCELERATION は  
 サーキットマップを用いず, レーシングカーの走行軌跡を  
 手がかりに進行方向の加速および進行方向とは垂直の方向  
 への加速度を素性として用いる.

既存データセット [3] を用いた実験より, 3つの提案素  
 性は実況生成問題およびプランニング問題に対し有益で  
 あることがわかった. 発話分類問題に対しては, 発話テキ  
 ストそのものの情報が強力であり, 生データや提案素性と  
 いったマルチモーダル情報を活用しても性能は向上しない  
 ことがわかった. 本稿で提案するような運動力学に基づく  
 計算は物理学分野で一般的である. 一方, 自然言語処理分  
 野での活用は稀である. 運動力学の知見による素性設計手  
 法は, レーシングゲーム実況のモデリングタスク以外にも  
 天気予報生成 [11] における雲の動きや, 市況コメント生  
 成 [1], [12] における株価の値動きを表現した折れ線グラ  
 フのモデリングなど, 物体や物事の動きを捉える必要のある  
 他の言語処理タスクにも汎用的に拡張できる可能性がある.

本稿の貢献は 1) 運動力学的な計算によりセンサーから  
 取得した生データに対しより深い解釈を与えた素性を提案  
 する, 2) 外部知識としてサーキットマップを用い, 時系列  
 数値データと外部知識を組み合わせる手法を提案する, 3)  
 実験より運動力学的な計算による素性が実況テキスト生成

およびプランニングにおいて効果的であることを示す。

## 2. 関連研究

従来、実況テキスト生成に関する研究はサッカー [6], [19], [20], 野球 [4], ゲーム録画映像 [3], [25] を対象に行われている。これらは生成ベースの手法 [4], [20], 抽出ベースの手法 [6] を用いるものに分類できる。本研究では前者に着目する。本稿では映像とともに視聴者に伝えられる実況を扱うが、テキストのみがユーザに提示されるテキスト速報を出力する設定 [6], [20] も存在する。

プランニング問題は自動言語生成における重要な問題設定として長らく取り組まれており、現在でも活発に研究されている [8], [15]。実況テキストを対象とするプランニングについては上田ら [28] の研究が存在する。

本研究で扱う発話分類問題は発話行為推定問題に近い。発話行為推定は古くは Switchboard コーパスなど電話応答音声の分析 [18] から始まり、メールスレッド [13] など他のテキストも対象とされている。本研究でのプランニングおよび発話分類には Ishigaki らの定義したラベルを用いる [3]。推定した発話ラベルは要約 [13] や発話生成 [24] といった下流タスクでの使用を想定しており、推定性能は下流タスクの性能に影響することから重要である。

外部知識の活用は固有表現抽出 [17], 関係抽出 [9] など多くの言語処理タスクで有効なことが報告されている。既存研究では知識ベースを外部知識として扱う設定が多く、テキストもしくは構造化データである。これらは埋め込み表現を得ることでモデルの入力として扱うことが可能である。一方、本研究で対象とするサーキットマップは二次元の座標位置の集合と捉えることができ、モデルへの入力としてどのように表現すべきか自明でない。人間の持つ常識知識をモデルに入力可能な形式に変換することも容易ではない。例えばサッカーや野球のルールを常識知識と見なす場合、パスやシュート、投球、攻撃や守備の交代などゲームの切れ目が明確である。このようなルールを人間の持つ知識と考えると、知識を素性として表現することは可能である。レーシングゲームの場合、ゲームの切れ目がなく連続的に進むことからレースのルールを素性として表現することが難しい。本研究では運動力学的な計算を用いて、外部知識や連続的に進むレースの状況を素性として表現する手法を提案する。

## 3. 問題設定

レーシング実況テキストとして公開データ [3] を用いる。<sup>\*1</sup> このデータセットには「Assetto Corsa」<sup>\*2</sup>の録画映像に対し、人間による実況音声とその文字起こしが付与さ

<sup>\*1</sup> <https://kirt.airc.aist.go.jp/corpus/en/RacingCommentary>

<sup>\*2</sup> Assetto Corsa は KUNOS Simulazioni 社の開発するレーシングゲームである。

れている。録画映像には空撮映像およびプレイヤー視点の2種類が存在する。さらに、Assetto Corsa API を用いて取得したゲーム内で用いられる速度やハンドル角度といった数値データも含まれる。従って、1つの発話に対し発話時刻を通じて時系列数値データが対応づいている。時系列数値データの種類を表1に示す。データセット中の発話のうち10レース分の571発話については、Ishigaki らの定義した発話分類ラベルが付与されている。発話分類ラベルの定義を表2に示す。以下に3つの下流タスクについて定式化する。

### 3.1 実況発話生成問題

実況発話生成問題は発話対象時刻が与えられた設定で、速度やハンドル角度を含む  $N$  種類の時系列数値データ  $R = \{r_1, \dots, r_N\}$  を用い、実況テキストを生成する。ここで、 $r_n$  は数値列を格納したベクトルである。時系列数値データの種類は既存研究 [3] と同様に表1に示すものを用いる。ほとんどの種類は連続値であるが、二値の値をとるものもある。その場合、0.0もしくは1.0の連続値として表現する。なお、 $r_n$  は発話対象時刻の過去10秒分を0.1秒ごとに記録した大きさ100のベクトルとする。

### 3.2 プランニング問題

プランニング問題 [27], [28] は、実況発話生成問題と同じく、時系列数値データ  $R$  が与えられ、次に何を発話すべきかを表現したラベルを出力する。発話ラベルには既存研究の定義によるラベル [3] を用いる。1つのラベルは着目するオブジェクトを表現する「対象サブラベル」と言及内容を表現する「内容サブラベル」の対として定義される。この問題は、複数のラベルの出力を許容するマルチラベル分類である。

### 3.3 発話分類問題

前述したプランニング問題は時系列数値データ  $R$  のみを入力として、ラベルを予測したのに対し、発話分類問題は分類対象発話テキスト  $T$  も入力とし、表2に示すラベルを出力する。この問題もマルチラベル分類である。

## 4. 手法

本節ではベースライン手法および提案手法について述べる。3つの問題に対するモデルの概要を図2に示す。それぞれのモデルでは同一のエンコーダを用い、時系列数値データを読み込む。発話分類問題では、このエンコーダに加え Lin らの提案する LSTM と Transformer を組み合わせたエンコーダ [10] で分類対象発話テキストを読み込む。ベースラインエンコーダは、入力データとして従来素性 [3] のみを用い、提案エンコーダは提案する3つの素性をそれぞれ用いる。以後、ベースラインエンコーダについてまず

構造化データの種類	値の例
現在の周回数 [0..]	1
レース中か否か	false
現在の周回のラップタイム (ms)	256
直前の周回のラップタイム (ms)	156164
現在のラップの進捗 [0, 1]	0.002780
ベストラップタイムとの差分	0.0
速度 (km/h)	177.693130
ハンドル角度 (rad)	-59.793526
高度, 緯度, 経度	(5.372770, 64.056038, 130.219971)
サーキット上の横方向位置 (L=-1, R=1)	-0.515301
理論上の最速コース取りとの距離 (m)	0.854022

表 1 時系列数値データの例 [3].

対象サブラベル	例
プレイヤーの車	ここは華麗な追い抜き, 決めていく
他の車	後ろの車に抜かれましたね.
すべての車	全車今一斉にスタート.
サーキット	本日のサーキット, ここラグナセカは長いストレートで有名です.
内容サブラベル	例
相対位置	プレイヤー現在 2 位.
絶対位置	青い車今第 2 カーブに差し掛かって, 他の車もそれに続く.
タイム	プレイヤー今ゴールラインを超えて 3 分 15 秒.
直前のイベント	このミスはタイムに響くかもしれない.
将来のイベント	難しいターン 15, 超えていけるか?
動き	プレイヤー, 長いストレートで追いついていく.
安定したレース	全車問題なく, レースは進んでいます.
特徴	今回のレース, 全車ボルシェ・マカン.
挨拶	はい, では実況を始めていきます.
反応	おお!

表 2 石垣ら [3] によるラベル定義. 1 つのラベルはサブラベルの組み合わせとして定義される.

説明し, その後提案素性の設計手法について述べる. その後, 3 つの下流タスクを解くモデルについて説明する.

#### 4.1 ベースラインエンコーダ

ベースラインエンコーダは  $R$  のみを入力とし, 各入力指標に対するベクトル  $\mathbf{r}_n$  を以下のように読み込む:

$$\mathbf{d}_n = \text{ReLU}(\mathbf{r}_n \mathbf{W}_{r_n} + b). \quad (1)$$

ここで,  $\mathbf{W}_{r_n}$  は 100 次元で表現された  $\mathbf{r}_n$  を埋め込み表現に変換する重み行列である. 活性化関数として ReLU を用いる.  $\mathbf{r}_1$  から  $\mathbf{r}_N$  をこのネットワークで埋め込み表現  $\mathbf{d}_1$  から  $\mathbf{d}_N$  に変換し, 以下のマルチヘッド自己注意機構によるエンコーダ MultiHeadAttn [23] で各指標間の関係も考慮した埋め込み表現集合を得る:

$$D = \text{MultiHeadAttn}(\mathbf{d}_1, \dots, \mathbf{d}_N). \quad (2)$$

ここで, MultiHeadAttn は各入力指標の埋め込み  $\mathbf{d}_n$  を指標間の関係も考慮する埋め込み  $D$  に変換する. 従って,  $D$  は  $N$  個の埋め込み表現が含まれる集合となり, 下流タスクを解くネットワークで用いる.

#### 4.2 提案エンコーダ

提案エンコーダは MultiHeadAttn の入力として, 提案素性である SECTIONS, CORNERS 及び ACCELERATION を追加する. 以下に 3 つの素性を順に説明する.

##### 4.2.1 SECTIONS: レーシングカーの位置とサーキットマップでの区間を対応づけ素性とする手法

空撮視点からの実況時に実況者はサーキットマップを外周知識として使用し, 必要に応じて「ターン 2」などのコーナーの番号を実況に含めることを許されている [3]. そこで, 本研究ではレーシングカーの走行している座標位置を外周知識であるサーキットマップの区間に割り当て素性とする手法を提案する. サーキットマップの例を図 3 に示す. 全てのマップは Assetto Corsa が提供するものを用いる.

Ishigaki ら [3] の分析によると, 「長いストレートで有名なムジェロサーキットが本日の会場です.」といったような, サーキットの特徴について述べる内容はレースの開始時刻周辺で言及される傾向がある. 一方, ラップタイムについてはレース終盤において言及されやすい. そこで, サーキットのスタートラインからゴールラインまでを 10 分割し, それを 2 周分異なる区間として表現することを考える. データセットに含まれるサーキットの種類数は 6 であるため, レーシングカーが現在走行する位置がどのサーキットのどの区間であるかを大きさ 120 ( $120 = 6 \times 10 \times 2$ ) の one-hot ベクトル  $\mathbf{r}_{sec}$  で表現する.  $\mathbf{r}_{sec}$  は以下の式で埋め込み表現  $\mathbf{d}_{sec}$  へと変換し, MultiHeadAttn への追加入力とし新たな  $D$  を得る:

$$\mathbf{d}_{sec} = \text{ReLU}(\mathbf{r}_{sec} \mathbf{W}_{sec} + b), \quad (3)$$

$$D = \text{MultiHeadAttn}(\mathbf{d}_1, \dots, \mathbf{d}_N, \mathbf{d}_{sec}). \quad (4)$$

これにより, サーキットマップを用い, サーキット場での区間も考慮したベクトル集合を得る. このエンコード手法を SECTIONS と呼ぶ.

##### 4.2.2 CORNERS: サーキットマップ上のコーナーを自動認識し素性とする手法

実況ではコーナーについて多く言及される. 従って, 実況者は SECTIONS のようにレーシングカーの走行位置とサーキットマップ上での区画を対応づけるだけでなく, レーシングカーの走行位置がコーナーであるかストレートであるかも認識していると仮定する. サーキットマップは座標位置の集合として表現されていることから, 各点についてコーナーであるかストレートであるかの明示的なアノテ

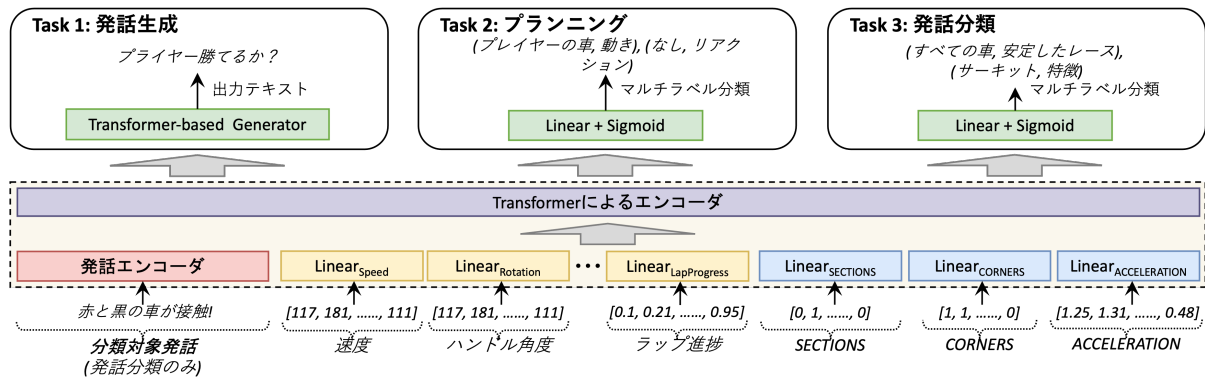


図 2 実況テキスト生成，プランニング及び発話分類問題のためのモデル。

ションが存在しない。そこで，コーナーか否かの二値を運動力学に基づく計算により自動付与を試みる。短い距離の間にコースの方向が大きく変化すればコーナーであるとみなすことができる。運動力学における曲率  $\kappa$  の値を用いて捉える手法を提案する。

コースは座標位置の集合で表現されており，隣接する座標位置ペアを  $(x_t, y_t)$  及び  $(x_{t-1}, y_{t-1})$  とする。ここで 2 点を結んだ時の方向を  $\phi_t = \arctan(\Delta y / \Delta x)$  として計算する。 $\Delta x$  と  $\Delta y$  はそれぞれ経度方向および緯度方向への移動距離である。さらに，2 つの時刻での方向の変位  $\Delta\phi = \phi_t - \phi_{t-1}$  も計算する。方向の変位に加え，2 点間の距離  $\Delta s = \sqrt{(\Delta x)^2 + (\Delta y)^2}$  を計算する。最終的に曲率は  $\kappa = \Delta\phi / \Delta s$  と計算される。曲率が閾値よりも大きければ，短い距離で大きく方向が変化したことになるためコーナーであると考えられる。曲率  $|\kappa| \geq 0.03$  の条件を満たす隣接点をコーナーとし，それ以外をストレートとする。閾値はコーナーについて明確な定義がないため設定が難しい。そこで，0.1 から順に値を大きくし，図 3 のように可視化した上で，著者のうち 2 人が目視し合議により 0.03 に決定した。提案手法によりアノテーションしたサーキットマップの例を図 3 に示す。この図で，青部分はコーナーであり，黒部分はストレートである。

過去 10 秒間についてレーシングカーの走行位置をサーキットマップ上の座標位置に対応づけ，その点がコーナーかストレートであるか判定し二値の素性ベクトル  $\mathbf{r}_{cor}$  とする。 $\mathbf{r}_{cor}$  を以下のように埋め込み表現  $\mathbf{d}_{cor}$  に変換した上でエンコードすることで，新たなベクトル集合  $D$  を得る：

$$\mathbf{d}_{cor} = \text{ReLU}(\mathbf{r}_{cor} \mathbf{W}_{cor} + b), \quad (5)$$

$$D = \text{MultiHeadAttn}(\mathbf{d}_1, \dots, \mathbf{d}_N, \mathbf{d}_{cor}). \quad (6)$$

このエンコード手法を CORNERS と呼ぶ。

#### 4.2.3 ACCELERATION: 加速度を用い加減速と回転を捉える素性

コーナーや，クラッシュといった重要な局面ではレーシングカーが回転したり，レーシングカーの速度が変化する。実況者はレーシングカーの回転や速度変化といった状況に

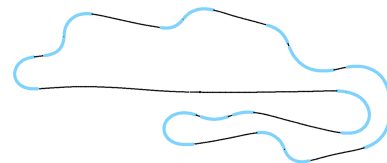


図 3 提案手法により自動アノテーションしたコーナーとストレート。青部分がコーナー，黒部分がストレート。

よって実況内容を変化させた方が良いことを知っている，と仮定する。運動力学における加速度はレーシングカーの速度や方向の変化を捉える基盤となる指標である。よって，加速度をこのような知識を捉える素性として用いる。具体的には，レーシングカーの過去の走行軌跡から加速度を 0.1 秒おきに 10 秒間分計算し，大きさ 100 の素性ベクトルを 2 つ作成する。1 つ目は進行方向への加速度を素性とし，レーシングカーの加減速を捉える素性  $\mathbf{r}_{accel}$  である。2 つ目は進行方向に対し垂直方向の加速度  $\mathbf{r}_{cent}$  を素性とする。この値はレーシングカーの回転しやすさと相関する。なお，この手法では走行軌跡を用い，前述した 2 手法とは異なり外部知識としてのサーキットマップは用いない。

図 4 に素性値計算の概要を示す。素性値は 4 つの手順で計算する：1) オレンジ色の矢印の始点で示すように，走行軌跡上の隣接する 2 点 ( $\mathbf{p}_{t_{i-1}}$  と  $\mathbf{p}_t$ ) について座標位置を得て位置ベクトルとする，2) 位置ベクトルを手がかりにオレンジの矢印で示すように速度ベクトル ( $\mathbf{v}_{t_{i-1}}$  と  $\mathbf{v}_t$ ) を得る，3) 速度ベクトルを用いて加速度ベクトル ( $\mathbf{a}_{t_i}$ ) を得る，4) 進行方向成分ベクトル ( $\mathbf{a}_{accel, t_i}$ ) と向心方向ベクトル  $\mathbf{a}_{cent, t_i}$  に成分分解。過去 10 秒間分の隣接点に対しこの手順を繰り返し， $\mathbf{r}_{accel}$  及び  $\mathbf{r}_{cent}$  を得る。

4 つの手順について具体的な計算式を示す。0.1 秒おきに隣接点の対応する位置ベクトル  $\mathbf{p}_{t_i}$  及び  $\mathbf{p}_{t_{i-1}}$  を得る。ここで， $t_i$  は素性値の計算対象となる時刻， $t_{i-1}$  は直前の時刻となる。従って，2 つの時刻の差分は  $\Delta t = t_i - t_{i-1} = 0.1$  となる。

これらの位置ベクトルから速度ベクトル  $\mathbf{v}_{t_i}$  及び加速度ベクトル  $\mathbf{a}_{t_i}$  を計算する：

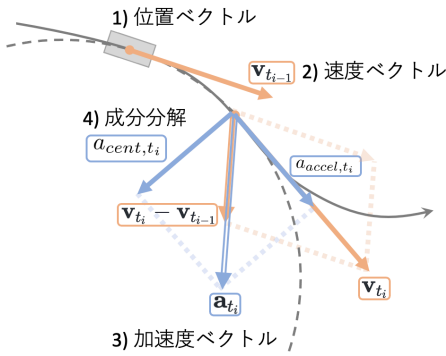


図 4 加減速と回転を捉える素性の計算手順.

$$\mathbf{v}_{t_i} = \frac{\mathbf{p}_{t_i} - \mathbf{p}_{t_{i-1}}}{\Delta t}, \quad (7)$$

$$\mathbf{a}_{t_i} = \frac{\mathbf{v}_{t_i} - \mathbf{v}_{t_{i-1}}}{\Delta t}. \quad (8)$$

$\mathbf{a}_{t_i}$  を進行方向  $a_{accel,t_i}$  と向心方向  $a_{cent,t_i}$  に成分分解する:

$$a_{accel,t_i} = \mathbf{a}_{t_i} \cdot \mathbf{e}_v, \quad (9)$$

$$a_{cent,t_i} = |\mathbf{a}_{t_i} \cdot \mathbf{e}_c|. \quad (10)$$

ここで,  $\mathbf{e}_v = \frac{\mathbf{v}_{t_i}}{|\mathbf{v}_{t_i}|}$  は進行方向の単位ベクトル,  $\mathbf{e}_c$  ( $\mathbf{e}_v \cdot \mathbf{e}_c = 0, \|\mathbf{e}_c\| = 1$ ) は向心方向の単位ベクトルで,  $|\mathbf{x}|$  は  $\mathbf{x}$  のノルムである. 0.1 秒おきに 10 秒間  $t_i$  を変化させながら  $a_{accel,t_i}$  を計算し, 大きさ 100 の素性ベクトル  $\mathbf{r}_{accel}$  を得る. 同様に  $a_{cent,t_i}$  も計算し,  $\mathbf{r}_{cent}$  を得る. 2 つの素性ベクトルはそれぞれ埋め込み表現に変換し, Transformer ベースのエンコーダで既存素性と共に読み込む:

$$\mathbf{d}_{accel} = \text{ReLU}(\mathbf{r}_{accel} \mathbf{W}_{accel} + b), \quad (11)$$

$$\mathbf{d}_{cent} = \text{ReLU}(\mathbf{r}_{cent} \mathbf{W}_{cent} + b), \quad (12)$$

$$D = \text{MultiHeadAttn}(\mathbf{d}_1, \dots, \mathbf{d}_N, \mathbf{d}_{accel}, \mathbf{d}_{cent}). \quad (13)$$

このエンコード手法を ACCELERATION と呼ぶ. 既存研究では速度変化や回転をセンサーデータから取得した速度などの生データからのみ捉えようとしている. 本研究では, 運動力学的な計算を用いて生データに対しより深い解釈を加える. この手法は外部知識としてのサーキットマップや特別なセンサーを用いない. GPS により取得した座標位置データのみを使用することから, 応用上の利点も大きい.

### 4.3 下流タスクのためのモデル

#### 実況発話生成モデル

実況発話生成は Transformer ベースの言語生成器としてモデリングする. このモデルは  $D$  及び直前に出力した語の埋め込み表現を受け取り, 次に出力する単語を繰り返し予測する. 具体的にはマルチヘッド注意機構ネットワークによる言語生成モデル [23] を用いる:

$$E = \text{DecoderAttn}(D, \text{Emb}(o_{j-1})), \quad (14)$$

$$o_j = \text{Softmax}(\text{Average}(E) \mathbf{W}_d). \quad (15)$$

ここで,  $\text{DecoderAttn}$  はエンコード済みのベクトル集合  $D$

と直前に出力した語の埋め込み表現を受け取り, 各入力ベクトルに対応する出力ベクトルの集合  $E$  を出力する.  $E$  の平均ベクトルを重み行列  $\mathbf{W}_d$  を用いて, 事前に定義した辞書サイズと同じの大きさのベクトルに変換する.  $\text{Softmax}$  関数を適用した後に得られる確率分布の各要素は, 辞書に含まれる各単語に対するスコアに対応する. スコア最大の語を出力するステップを, 文の終わりを示す特別なトークンを出力するか, 最大出力長に達するまで繰り返す. 重み行列は交差エントロピー損失を用いて学習する.

#### プランニングモデル

プランニングモデルはエンコード済みのベクトル集合  $D$  を受け取り<sup>\*3</sup>, ラベルを予測する:  $\mathbf{p}_{plan} = \text{sigmoid}(\mathbf{W}_p \text{Average}(D))$ . ここで  $\mathbf{W}_p$  はエンコード済みのベクトル  $D$  の平均ベクトルを候補ラベル数と同じサイズに変換する重み行列である.  $\text{sigmoid}$  関数により各要素はそれぞれ確率値に変換され, この値が各ラベルに対するスコアに対応する. プランニングは複数のラベル付与を許容するマルチラベル設定であり,  $\text{Softmax}$  関数ではなく  $\text{sigmoid}$  関数を用いる. 閾値  $\delta$  以上のスコアを得たラベルを出力する.  $\delta$  は 0.1 から 1.0 の間で 0.1 刻みに調整し, 開発セットでの F 値が最大となる値を採用する.  $\delta$  以上のスコアを得るラベルが存在しない場合は何も出力しない<sup>\*4</sup>. このネットワークは二値交差エントロピー損失を用いて学習する.

#### 発話分類モデル

入力の発話テキストはまず LSTM [2] で読み込み, その後さらに自己注意機構によるエンコーダ [10] で埋め込み表現  $\mathbf{t}$  に変換する.  $\mathbf{t}$  は時系列数値データのエンコード済みベクトルの平均  $\text{Average}(D)$  と結合された後, スコア分布へと変換される:  $\mathbf{p}_{cls} = \text{sigmoid}(\mathbf{W}_c[\text{Average}(D); \mathbf{t}])$ . ここで,  $[\cdot]$  は結合ベクトルを表す.  $\mathbf{W}_c$  は候補ラベル数と同じ大きさのベクトルを得るための重み行列である.  $\text{sigmoid}$  を適用することで, 各スコアを確率値に変換する. 最終的に  $\mathbf{p}_{cls}$  はラベル候補数と同じ大きさのベクトルで, 各要素が対応するラベルに対するスコアとなる. スコアが閾値  $\delta$  よりも大きいラベルを 1 つ以上出力しうるマルチラベル設定となる.  $\delta$  は開発セットでの F 値が最大となる値に 0.1 から 1.0 の間で 0.1 刻みに調整する. このネットワークは二値交差エントロピー損失を用いて学習する.

## 5. 実験

データセット, パラメータ, 比較モデルと評価指標について述べる.

<sup>\*3</sup> 実際の実装では実況時に使用した映像が「空撮視点」か「プレイヤー視点」の映像かを示す 2 次元の埋め込み表現も入力している.

<sup>\*4</sup> ただし, そのような事例は評価セットには存在しない.

## 5.1 データセットとパラメータ

**実況発話生成:** 実験には Ishigaki ら [3] の公開データセットを用いる。このデータセットには 2,400 レースに対し付与された 129,226 発話の実況テキストが含まれる。空撮視点及びプレイヤー視点の映像を用いた実況が含まれるが、実況者がサーキットマップを使用するよう指示されたのは空撮視点のみである。よって、空撮視点の実況 1,084 レース分を実験に用いる。これらのレースを 3:1:1 の割合で分割し、それぞれ学習、開発、評価のために用いる。トークン化には MeCab [7] を用い、辞書には IPADic を使用した。トークンの埋め込みは 40 次元に設定した。最適化には Adam [5] を用い、初期学習率は 0.001 に設定した。バッチサイズは 100 とし、100 エポック学習を行い、開発セットでの BLEU が最も高くなるモデルを選択し評価した。

**プランニングと分類:** Ishigaki らが分析に用いた 571 実況発話を実験に用いる。これらの発話には発話ラベルが付与されている。訓練データとして 80%、評価データとして 20% を用い、5 分割交差検定を行う。各分割で訓練データとして用いるデータのうち 25% を開発データとしてハイパーパラメータの調整に使用し、残りのデータをモデルの訓練に用いた。また、対象発話テキスト及び直前の発話テキストをエンコードする際の、単語埋め込み表現の次元は 300 とした。MeCab の辞書には IPADic を用いた。LSTM の隠れ状態の次元は 64 とし、ドロップアウト率は 0.2 に設定して学習させた。自己注意機構の深さ  $r$  は 3 に設定した。自己注意機構の重み  $A$  の推定に用いる中間表現の次元数は 16 とした。また、テキストデータを単語ベースで埋め込み表現へと変換する際には、その初期値として形態素を埋め込み表現に変換するための重み行列はランダムに初期化し、学習時に更新した。異なる初期値で 5 回ずつ実験を行いその平均値を評価した。ネットワークの最適化は Adam [5] の初期学習率を 0.0001 に設定し行った。学習時のバッチサイズは 16 とした。また、学習時には二値交差エントロピー損失を用い、20 エポックの間開発セットでの損失関数の値が改善しない場合に学習を終了した。

## 5.2 比較モデルと評価指標

既存素性 [3] のみを用いるモデルをベースラインとし、提案素性も用いるモデルと比較する。ベースラインを BASE とし、提案素性を用いるモデルをそれぞれ BASE+SECTIONS, BASE+CORNERS 及び BASE+ACCELERATION とし比較する。プランニング問題と分類問題では、追加のベースラインとして学習データ内で頻出する上位  $k$  ラベルを出力するモデルとも比較する。1 発話あたりに付与されたラベル数の平均は 1.4 であることから、 $k$  は 1 及び 2 とする。また、強いベースラインとして  $k$  を正解ラベル数とした手法とも比較する。

発話テキスト生成問題の評価には既存研究 [3] と同様に

	BLEU
BASE	5.49
BASE + SECTIONS	<b>6.49*</b>
BASE + CORNERS	6.44*
BASE + ACCELERATION	5.85*
BASE + All	6.06*

表 3 実況発話生成の BLEU スコア。提案素性を加えたモデルはベースライン (BASE) よりも高い値を示した。\* BASE との差が並べ替え検定により統計的に有意な手法 ( $p < 0.05$ )。

	Prec.	Rec.	F1
k-頻出 ( $k = 1$ )	.351	.243	.287
k-頻出 ( $k = 2$ )	.297	.412	.345
k-頻出 ( $k =$ 正解ラベル数)	<b>.360</b>	.360	.360
BASE	.331	.398	.354 ±.006
BASE + SECTIONS	.305	<b>.470</b>	.363 ±.006
BASE + CORNERS	.318	.455	<b>.367 ±.011</b>
BASE + ACCELERATION	.315	.440	.363 ±.008
BASE + All	.314	.445	.363 ±.005

表 4 プランニングタスクの結果。提案素性により性能が向上することを確認した。

	Prec.	Rec.	F1
k-頻出 ( $k = 1$ )	.351	.243	.287
k-頻出 ( $k = 2$ )	.297	.412	.345
k-頻出 ( $k =$ 正解ラベル数)	.360	.360	.360
BASE	<b>.656</b>	.520	<b>.576</b>
BASE + SECTIONS	.627	<b>.524</b>	.568
BASE + CORNERS	.627	.516	.561
BASE + ACCELERATION	.622	.508	.556
BASE + All	.629	.499	.553

表 5 発話分類の結果。この問題では、発話テキストそのものの情報が強力であり、数値列データは性能向上に寄与しない。

BLEU [14] を用いる。また、コーナーの番号についての程度正しく言及しているか人手で確認し、精度を計算する。プランニングおよび分類は適合率、再現率、F 値を用いる。

## 6. 結果

3 つの下流タスクでの性能を報告する。

### 6.1 自動定量評価

**実況発話テキスト生成:** 表 3 に BLEU スコアを示す。BASE は 5.49 を示した。提案素性である SECTION, CORNERS 及び ACCELERATION を追加することで、BLEU スコアはそれぞれ 6.49, 6.44 及び 5.85 に向上した。これらの値は BASE との差が統計的に有意であることを並べ替え検定により確認した ( $p < 0.05$ )。SECTIONS 及び CORNERS はこの問題において ACCELERATION よりも良い性能を示すことから、外部知識としてサーキットマップを用いることは効果的であることがわかる。一方で、ACCELERATION にはサーキットマップを必要としないという実用上の利点がある。全て

の素性を組み合わせたモデル BASE+ALL の BLEU は 6.06 であり、各素性を単体で追加したモデルよりも良い値を得るわけではない。効果的な組み合わせについては更なる検証が必要である。

**プランニング:** 表 4 に適合率, 再現率, F 値を示す。BASE は F 値において 0.357 を示した。この値は、正解ラベル数を出力する  $k$  頻出ベースラインの 0.360 よりも低い値である。よって、 $k$  頻出ベースラインは強いことがわかる。SECTIONS, CORNERS 及び ACCELERATION を追加すると、ベースラインの 0.357 からそれぞれ F 値が 0.363, 0.367, 0.363 に上昇する。発話テキスト生成とは異なり、ACCELERATION が最も性能向上に貢献する。SECTIONS は発話テキスト生成とは異なり、性能向上に大きく寄与しない。この理由は、学習データが非常に小さくサーキットマップを 20 分割した際にかく区間に対応する学習事例数が小さいためと思われる。ACCELERATION はサーキットマップを必要とせず、特定のサーキットに依存しない汎用的な素性が獲得できる利点がある。

**発話分類:** 表 5 に発話分類問題の適合率, 再現率, F 値を示す。BASE は F 値において 0.576 を示し、この値は  $k$ -頻出ベースラインよりも 0.2 ポイント以上高い。一方、提案素性の SECTIONS, CORNERS 及び ACCELERATION を追加しても、性能はそれぞれ 0.568, 0.561, 0.556 と性能向上しない。この問題では発話テキスト自体が発話分類のために十分な情報を持っていると考えられ、この結果自体妥当である。よって、素性が増えてネットワークが複雑になるに従い性能が劣化するものと思われる。

## 6.2 正しさに関する人手分析

BLEU は生成テキストの正しさを正確に評価できない場合がある [22]。そこで、BLEU の評価に加え人間の目視による確認を行い生成テキストの品質について分析する。分析では、コーナーについての言及の正しさに着目する。具体的には「ターン 2」などのコーナー番号を正しく含めた発話の割合を正解率と定義する。すなわち、 $C$  をコーナー番号を正しく含められた発話の数とし、 $T$  を評価対象とした発話の数とし、正解率を  $\frac{C}{T}$  として計算する。評価対象として、BASE と BASE+SECTIONS の生成テキストが異なる事例からランダムに 50 事例抽出した。BASE 及び BASE+SECTIONS を人間による実況発話と比較し、コーナー番号が正しいか否かを人手で付与した。テキストのみでの判定が難しい場合には映像も確認した。

正解率を表 6 に示す。BASE の正解率が 0.40 であるのに対し、BASE+SECTIONS では 0.70 と改善する。実際の生成テキストの例を表 7 に示す。例 1 に示すように BASE では「ターン 11」を「ターン 12」と出力するといったエラーが散見される。BASE+SECTIONS ではこのようなエラーの一部に改善が見られた。例 2 は評価の人手コストが大きな

	BASE	BASE+SECTIONS
正しさ	0.40 (20/50)	0.70 (35/50)

表 6 正しさについての人手評価。

例 1	
GOLD	次, ターン 11
BASE	ターン 12 抜けた.
BASE + CORNERS	ターン 11.
例 2	
GOLD	今, ターン 2 に入ります.
BASE	第 3 コーナーを進んでいく.
BASE + CORNERS	ターン 1 抜ける.

表 7 生成例。

事例である。人間による発話は「ターン 2 に突入。」で、BASE+SECTIONS は「ターン 1 は無事抜けました。」である。自動生成した発話が正しいか否かはターン 1 とターン 2 が連続しているか否かによる。すなわち、2 つのターンが連続していれば「ターン 1 を抜ける」ことは「ターン 2 への突入」は同じ状況であり正解となるが、連続するターンでない場合は不正解である。このような事例の評価を行うためには、映像の確認とサーキットマップの目視が必要となり、評価にかかるコストが高い。マルチモーダルな設定における効率的な評価指標については今後の研究課題とする。

## 7. 終わりに

本稿ではレーシングゲーム実況テキストモデリングのための素性を 3 つ提案した。より良いモデリングのために、外部知識としてサーキットマップを活用し生データに対してより深い解釈を加えることで性能が向上することを示した。また、運動力学的な計算を用いることで性能が向上することがわかった。運動力学的な計算による素性設計手法は、物体の動きを捉える必要のある天気予報生成など他の言語処理タスクへの応用可能性がある。このような素性設計を用いずともニューラルネットワークが速度などの生データから自動学習できることが理想ではある。しかし、現実的には学習データ、計算資源の制約があるため、提案素性のような解釈を加えた素性の設計が重要である。このような素性設計手法は効率的に高品質な言語処理モデルの学習を可能にする一つの手段であり、近年注目されている「efficient NLP」の方向性にも寄与できる可能性がある。知識と NLP の統合という観点から本研究は、従来から用いられてきた構造化データだけでなく、サーキットマップのような二次元データや人間に持つ知識といった構造化しにくいデータを、運動力学など他分野の知見を用いることで柔軟に NLP モデルに統合できる可能性がある。



## 謝辞

この成果は、国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) の助成事業 (JPNP20006) の結果得られたものである。産総研の AI 橋渡しクラウド (ABCI) を利用し実験を行った。

## 参考文献

- [1] Aoki, T., Miyazawa, A., Ishigaki, T., Goshima, K., Aoki, K., Kobayashi, I., Takamura, H. and Miyao, Y.: Generating Market Comments Referring to External Resources, *Proceedings of the 11th International Conference on Natural Language Generation (INLG2018)*, Tilburg University, The Netherlands, Association for Computational Linguistics, pp. 135–139 (online), DOI: 10.18653/v1/W18-6515 (2018).
- [2] Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, *Neural Computation*, Vol. 9, No. 8, pp. 1735–1780 (online), DOI: 10.1162/neco.1997.9.8.1735 (1997).
- [3] Ishigaki, T., Topic, G., Hamazono, Y., Noji, H., Kobayashi, I., Miyao, Y. and Takamura, H.: Generating Racing Game Commentary from Vision, Language, and Structured Data, *Proceedings of the 14th International Conference on Natural Language Generation (INLG2021)*, Aberdeen, Scotland, UK, Association for Computational Linguistics, pp. 103–113 (online), available from <https://aclanthology.org/2021.inlg-1.11> (2021).
- [4] Kim, B. J. and Choi, Y. S.: Automatic Baseball Commentary Generation Using Deep Learning, pp. 1056–1065 (online), available from <https://doi.org/10.1145/3341105.3374063> (2020).
- [5] Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization (2015).
- [6] Kubo, M., Sasano, R., Takamura, H. and Okumura, M.: Generating Live Sports Updates from Twitter by Finding Good Reporters, *2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, Vol. 1, No. 1, pp. 527–534 (online), DOI: 10.1109/WI-IAT.2013.74 (2013).
- [7] Kudo, T., Yamamoto, K. and Matsumoto, Y.: Applying Conditional Random Fields to Japanese Morphological Analysis, *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, Barcelona, Spain, Association for Computational Linguistics, pp. 230–237 (online), available from <https://aclanthology.org/W04-3230> (2004).
- [8] Kukich, K.: Design of a Knowledge-Based Report Generator, *Proceedings of 21st Annual Meeting of the Association for Computational Linguistics (ACL1983)*, pp. 145–150 (online), DOI: 10.3115/981311.981340 (1983).
- [9] Li, P., Mao, K., Yang, X. and Li, Q.: Improving Relation Extraction with Knowledge-attention, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP2019)*, Hong Kong, China, Association for Computational Linguistics, pp. 229–239 (online), DOI: 10.18653/v1/D19-1022 (2019).
- [10] Lin, Z., Feng, M., dos Santos, C. N., Yu, M., Xiang, B., Zhou, B. and Bengio, Y.: A Structured Self-Attentive Sentence Embedding, *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24–26, 2017, Conference Track Proceedings*, OpenReview.net, (online), available from [https://openreview.net/forum?id=BJC\\_jUqxe](https://openreview.net/forum?id=BJC_jUqxe) (2017).
- [11] Murakami, S., Tanaka, S., Hangyo, M., Kamigaito, H., Funakoshi, K., Takamura, H. and Okumura, M.: Generating Weather Comments from Meteorological Simulations, *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, Online, Association for Computational Linguistics, pp. 1462–1473 (online), DOI: 10.18653/v1/2021.eacl-main.125 (2021).
- [12] Murakami, S., Watanabe, A., Miyazawa, A., Goshima, K., Yanase, T., Takamura, H. and Miyao, Y.: Learning to Generate Market Comments from Stock Prices, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (ACL2017)*, Vancouver, Canada, Association for Computational Linguistics, pp. 1374–1384 (online), DOI: 10.18653/v1/P17-1126 (2017).
- [13] Oya, T. and Carenini, G.: Extractive Summarization and Dialogue Act Modeling on Email Threads: An Integrated Probabilistic Approach, *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, Philadelphia, PA, U.S.A., Association for Computational Linguistics, pp. 133–140 (online), DOI: 10.3115/v1/W14-4318 (2014).
- [14] Papineni, K., Roukos, S., Ward, T. and Zhu, W.-J.: Bleu: a Method for Automatic Evaluation of Machine Translation, *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, Pennsylvania, USA, Association for Computational Linguistics, pp. 311–318 (online), DOI: 10.3115/1073083.1073135 (2002).
- [15] Puduppully, R., Dong, L. and Lapata, M.: Data-to-Text Generation with Content Selection and Planning, *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI2019)*, pp. 6908–6915 (online), DOI: 10.1609/aaai.v33i01.33016908 (2019).
- [16] Schaffrath, M.: Mehr als 1:0! Bedeutung des Live-Kommentars bei Fußballübertragungen – eine explorative Fallstudie [More than 1:0! The importance of live commentary on football matches – An exploratory case study], *Medien und Kommunikationswissenschaft*, Vol. 51 (2003).
- [17] Seyler, D., Dembelova, T., Del Corro, L., Hoffart, J. and Weikum, G.: A Study of the Importance of External Knowledge in the Named Entity Recognition Task, *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (ACL2018)*, Melbourne, Australia, Association for Computational Linguistics, pp. 241–246 (online), DOI: 10.18653/v1/P18-2039 (2018).
- [18] Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Van Ess-Dykema, C. and Meteer, M.: Dialogue act modeling for automatic tagging and recognition of conversational speech, *Computational Linguistics*, Vol. 26, No. 3, pp. 339–374 (online), available from <https://aclanthology.org/J00-3003> (2000).
- [19] Tanaka-Ishii, K., Hasida, K. and Noda, I.: Reactive Content Selection in the Generation of Real-time Soccer Commentary, *Proceedings of the 17th International Conference on Computational Linguistics (COLING1998)*, (online), available from <https://aclanthology.org/C98-2204> (1998).

- [20] Taniguchi, Y., Feng, Y., Takamura, H. and Okumura, M.: Generating Live Soccer-Match Commentary from Play Data, *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI2019)*, Vol. 33, No. 1, pp. 7096–7103 (online), DOI: 10.1609/aaai.v33i01.33017096 (2019).
- [21] Twitchell, D. P., Adkins, M., Nunamaker, J. F. and Burgoon, J. K.: Using speech act theory to model conversations for automated classification and retrieval, *Proceedings of the International Working Conference Language Action Perspective Communication Modelling (LAP2004)*, pp. 121–130 (2004).
- [22] Uehara, Y., Ishigaki, T., Aoki, K., Noji, H., Goshima, K., Kobayashi, I., Takamura, H. and Miyao, Y.: Learning with Contrastive Examples for Data-to-Text Generation, *Proceedings of the 28th International Conference on Computational Linguistics*, Barcelona, Spain (Online), International Committee on Computational Linguistics, pp. 2352–2362 (online), DOI: 10.18653/v1/2020.coling-main.213 (2020).
- [23] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u. and Polosukhin, I.: Attention is All you Need, *Advances in Neural Information Processing Systems* (Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S. and Garnett, R., eds.), Vol. 30, Curran Associates, Inc. (2017).
- [24] Wang, K., Tian, J., Wang, R., Quan, X. and Yu, J.: Multi-Domain Dialogue Acts and Response Co-Generation, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL2020)*, Online, Association for Computational Linguistics, pp. 7125–7134 (online), DOI: 10.18653/v1/2020.acl-main.638 (2020).
- [25] Xu, J. H., Fang, Z., Chen, Q., Ohno, S. and Paliyawan, P.: Fighting Game Commentator with Pitch and Loudness Adjustment Utilizing Highlight Cues, *2021 IEEE 10th Global Conference on Consumer Electronics (GCCE)*, pp. 366–370 (online), DOI: 10.1109/GCCE53005.2021.9621827 (2021).
- [26] Zhang, R., Li, W., Gao, D. and Ouyang, Y.: Automatic Twitter Topic Summarization With Speech Acts, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 21, No. 3, pp. 649–658 (online), DOI: 10.1109/TASL.2012.2229984 (2013).
- [27] 上田佳祐, 石垣達也, 小林一郎, 宮尾祐介, 高村大也: 実況における発話ラベル予測, 情報処理学会第 253 回自然言語処理研究会 (NL253), No. 1, pp. 1–6 (2021).
- [28] 上田佳祐, 石垣達也, 小林一郎, 宮尾祐介, 高村大也: 実況発話ラベル予測モデルにおける状況認識素性の活用, 言語処理学会第 28 回年次大会発表論文集, pp. 1371–1375 (2022).