Vol.2021-NL-251 No.1
Vol.2021-SLP-139 No.1
2021/12/1

実況における発話ラベル予測

上田 佳祐[1,2,a)]　石垣 達也[1,b)]　小林 一郎[1,3,c)]　宮尾 祐介[1,2,d)]　高村 大也[1,e)]

概要：本稿では，実況テキストにおける発話ラベル予測について述べる．実況の一例として，レーシングゲーム実況に着目する．例えば，「ターン 6，ここで追い抜いていけるか」という実況発話においては「サーキット場での位置」や「未来のイベントに関する情報」が述べられている．本研究ではこのような言及対象や発話内容に関する情報を，発話ラベルとして表現し，1) 分類対象発話テキストが与えられラベルを予測する設定，2) 発話テキストは与えられず，その時点までの文脈情報やレース状況を主に用いて次の発話のラベルを予測する 2 つのタスクを扱う．前者は従来の発話行為推定などの言語理解の問題に関連し，後者は言語生成におけるプランニングの問題と関連する．提案タスクでは，テキスト入力だけでなく，時間推移やレース映像を撮影するカメラの視点など複数のモダリティにより表現された情報も用いラベル予測を行う点が従来研究とは異なり，本研究では複数モダリティを活用するいくつかのモデルを比較する．実験より，ニューラルネットワークによる手法は最頻出ラベルを予測するたベースライン手法と比較し，特に分類対象発話テキストが与えられる設定において，ベースライン手法よりも高い F 値を得ることを確認した．一方，複数モダリティを考慮するモデルや発話テキストが与えられない設定においては単純なベースラインよりも性能が低く，提案課題が挑戦的な課題であることが分かった．

キーワード：発話タイプ同定，プランニング，マルチモーダル情報，言語生成

Utterance Label Prediction in Live Commentary

Keisuke Ueda[1,2,a)]　Tatsuya Ishigaki[1,b)]　Ichiro Kobayashi[1,3,c)]　Yusuke Miyao[1,2,d)]
Hiroya Takamura[1,e)]

1. はじめに

スポーツやゲーム映像に付与される実況は，実況者がイベントを認識し適切な言及対象や発話内容をリアルタイムに決定した上で，発話として表層化される．視聴者は映像を視聴しながら実況を聞くことで，よりイベントを理解し，映像を楽しむことが出来る．本研究では実況の一例として表 1 に示すような，レーシングゲーム映像に対する実況に着目する．レーシングゲーム実況において，実況者は各時刻において，「サーキット」や「プレイヤーの車」といった言及対象を決定し，「動き」や「(車両間の) 相対位置」といった具体的な言及内容を決定し発話を構成する．

本研究では，言及対象と言及内容のペアを 1 つのラベルと考え，これらを予測するタスクに取り組む．具体的には 2 つの発話ラベル予測タスクを考える; 1) 発話テキストを主な入力とし言及対象及び言及内容ラベルを同定するタスク，2) 過去の発話やレース状況を主な入力とし，次の言及対象と発話内容を同定するタスクである．前者を**対象発話ラベル予測タスク**，後者を**未来発話ラベル予測タスク**と呼ぶ．対象発話ラベル予測タスクでは，図 1 に示すような「プレイヤーは青のポルシェ」といった発話テキストを入力し，「プレイヤーの車/特徴」といったラベルを予測する．これは発話行為推定 [13] などの言語理解タスクの一つと捉えることが出来る．正しく発話ラベルを予測することで，

[1] 産業技術総合研究所
[2] 東京大学
[3] お茶の水女子大学
[a)] ueda.keisuke@aist.go.jp
[b)] ishigaki.tatsuya@aist.go.jp
[c)] koba@is.ocha.ac.jp
[d)] yusuke@is.s.u-tokyo.ac.jp
[e)] takamura.hiroya@aist.go.jp

時刻	発話テキスト (空撮視点)	ラベル
00:01	本日は長いストレートが有名なラグナセカのレースです.	サーキット/特徴
00:10	プレイヤーは青のポルシェ.	プレイヤーの車特徴
00:13	今全車一斉にスタート.	すべての車/動き
時刻	発話テキスト (ドライバー視点)	ラベル
00:01	今回は 5 番手からスタート.	プレイヤーの車/相対位置
00:10	華麗なスタートを決めていきたい.	プレイヤーの車/未来のプレー
00:13	あぁ追いつけない..	プレイヤーの車/過去のプレー

表 1 レーシング実況発話とその発話ラベルの例．先頭 3 発話は空撮映像に対し実況者が付与した実況，末尾 3 発話はドライバー視点の映像に対しドライバー自身が付与した実況である．

例えば，要約 [10] などの下位タスクへの応用可能性がある．未来発話ラベル予測タスクでは，テキストを与えず前の文脈テキストである「... ラグナセカのレースです.」や発話時刻などを入力とし，次の発話ラベルを予測する．これは過去の文脈を正しく理解し，次の発話内容を決定する言語生成研究におけるプランニング問題と捉えることができ，高品質な言語生成への応用を想定している．

本研究が対象とするレーシングゲームの実況においては，実況テキストそのもの以外にも，実況者の観る映像の視点や，レースの時間的な進捗の変化も，発話ラベル予測の手がかりとなりうる．映像の視点は具体的には，ゲーム内の仮想的なヘリコプターから撮影した映像からの視点である「空撮視点」及びプレイヤーが運転する車の背後に設置された仮想的なカメラから撮影される「プレイヤー視点」を考える．視点が異なると，表 1 の例のように着目する対象や言及内容に違いが生まれる．時間は，レース全体の中で発話が発せられる時刻を表す．Ishigaki ら [3] の分析によれば，ゲームの開始直後ではサーキットの特徴に関する言及が多く発せられるのに対し，終盤間際ではタイムに関する発話が増加するといった特徴がみられ，発話ラベル推定においても活用できる．本研究では発話テキストに加えて，このような視点や時間など補助的なデータも用い，発話ラベル予測を行うマルチモーダルな設定を考える．

本稿では 2 つの発話ラベル予測問題に対し，ベースライン手法とニューラルネットワークを用いた手法を適用した結果を報告する．レーシングゲーム映像に対する実況データを対象としたデータセットを用いた実験より，ニューラルネットワークを用いる手法は頻出するラベルを出力するベースラインよりも，対象発話ラベル予測タスクにおいて良い性能を示すことがわかった．一方，視点や時間などの補助的なデータを考慮する手法では，考慮しない手法と同等あるいはそれよりも劣化する結果を得ており，複数モダリティの組み合わせモデルの設計が難しく，挑戦的な課題であることを報告する．

2. 関連研究

実況テキストを対象とした研究は，主に自動言語生成の分野において，サッカー [7], [14], [15]，野球 [5]，ゲーム映像 [3] などいくつかのスポーツに対し行われている．本研究では実況の一例として，レーシングゲーム映像に対する実況データ [3] に着目する．このデータには，実況者が 2 つの異なる視点から実況したデータが含まれる．1 つ目はヘリコプターからレースの状況を撮影した空撮映像を見ながら「空撮視点」から実況をしたデータ，2 つ目はプレイヤーの運転する車の背後に設置されたカメラで撮影された映像を見ながらドライバーになったつもりで実況する「ドライバー視点」のデータであり，それぞれ実況の特徴が異なる．

本稿における対象発話ラベル予測タスクは，発話行為推定タスクと同様に発話に対するラベル予測である．古典的には Switchbord コーパスに含まれる電話応対対話に対し発話ラベルを予測する研究 [13] から始まり，メールスレッド [10] などへと対象が広がっている．このような問題は自然言語を計算機が理解するという学術的な関心だけでなく，例えば，メールスレッドの要約 [10] や応答生成 [16] などの応用システムにも活用される．発話行為推定は，「質問」「意見」といった言語行為論 [1], [12] に基づくラベル設計を用いるのが一般的であるが，レーシングゲーム実況に適用可能な発話ラベルとしては，既存研究 [3] で実況テキストの特徴分析をするためのラベルが提案されており，本研究ではそのラベル設計を用いる．一方，本稿における未来発話ラベル予測タスクは，「次に何について言及するか」を決定する従来の言語生成研究におけるプランニング [8], [11] の問題と捉えることができ，将来的には言語生成問題への応用を想定している．

発話ラベル予測問題は，分類問題もしくは系列タグ付け問題として定式化され，近年はニューラルネットワークによる手法が主に用いられる．本研究では発話ラベル予測を教師あり学習による分類問題として扱う．入力の観点からはテキストのみを扱う手法 [4]，映像や発話音声を用いる手法 [9] が存在する．本研究では，テキストに加え，実況に対して特徴的な情報として実況者が用いる映像の視点や時間など補助的なデータを組み合わせたマルチモーダルな入力を想定する．

3. 手法

本節では発話ラベルの設計とラベル推定手法について述べる．

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越していく.
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，全車ポルシェ・マカン.
挨拶	はい，では実況を始めていきます.
反応	おぉ*!*

表 2 サブラベルの一覧．ラベルは 2 つのサブラベルの組み合わせとして定義される．ラベルは Ishigaki ら [3] の定義に従う．

3.1 発話ラベル

予測するラベルには，既存研究 [3] においてレーシングゲーム実況の特徴分析に用いたラベル定義を採用する．ラベル定義と発話例を表 2 に示す．この定義において，1 つのラベルは 2 種類のサブラベルの組み合わせとして表現される．1 つ目のサブラベルは言及対象を表し，「プレイヤーの車」、「他の車」、「すべての車」及び「サーキット」のいずれかの値をとる．例えば「全車今一斉にスタート」という発話には「すべての車」というサブラベルが付与される．2 つ目のサブラベルは言及内容を表す．これは「タイム」、「車の動き」、「挨拶」など 10 種類の値を含む．例えば，「プレイヤー現在，タイム 3 分 15 秒.」という発話であれば，「タイム」のサブラベルが付与される．これらのサブラベルを 1 つのラベルとすることから，この発話には「プレイヤーの車/タイム」のラベルが付与される．1 つの発話に複数の言及対象や言及内容が含まれることがあるため，1 つの発話に対し 40 のラベル候補から複数のラベルが付与される，マルチラベル設定となる．

3.2 予測モデル

次に実験に用いた予測モデルについて述べる．本実験では 2 つのタスクに対して，共通の予測モデルを用いる．図 1 に示す提案手法は，入力データをそれぞれ異なるエンコーダで読み込み，その後それらを結合するモデルを採用する．対象発話ラベル予測タスクにおいては，予測モデルへの入力として対象発話テキスト，さらに視点情報や時間情報などの補助的なデータを与える．一方で未来発話ラベル予測タスクにおいては，対象発話テキストを入力として与えず，補助的なデータ及び直前の発話テキストのみを与える．

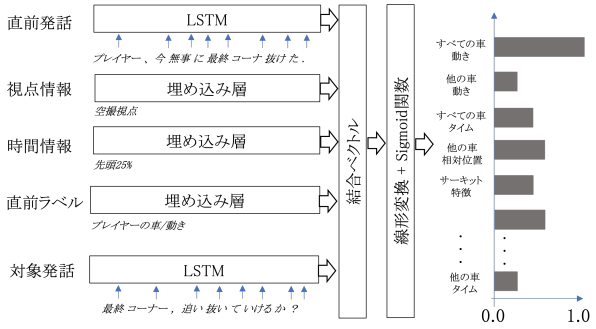

図 1 対象発話を与える設定の予測モデル．次の発話ラベルを予測する設定では，このモデルから「対象発話」が除去される．

まず，テキストデータは単語列で表現する．すなわち，対象発話テキスト $T = \{w_1, .., w_u\}$，直前の発話テキスト $T' = \{w'_1, .., w'_{u'}\}$ となる．視点情報データ D_1，時間情報データ D_2，直前の発話ラベル D_3 はそれぞれ別のエンコーダによって分散表現に変換し，テキストデータ及びそれら補助的なデータの分散表現を結合してエンコーダの最終出力とする．ここで，視点情報データ D_1 は空撮視点，ドライバー視点のいずれか，時間情報データ D_2 は 1 レース時間を 4 分割した区間のうち，発話時刻の区間を表す．また，対象発話がレース全体の最初の発話である場合は，直前の発話ラベルは与えられない．

対象発話テキスト T については，MeCab により分かち書きをした後に，各トークンに対応する埋め込み表現へと変換され，単語ベースの双方向 LSTM [2] で読み込む:

$$\overrightarrow{\mathbf{h}_i} = \mathtt{LSTM}(\overrightarrow{\mathbf{h}_{i-1}}, \mathtt{emb}(w_i)), \quad (1)$$

$$\overleftarrow{\mathbf{h}_i} = \mathtt{LSTM}(\overleftarrow{\mathbf{h}_{i+1}}, \mathtt{emb}(w_i)). \quad (2)$$

ここで，emb は対象の単語の埋め込み表現を返す関数である．さらに，ベクトル $\overrightarrow{\mathbf{h}_i}, \overleftarrow{\mathbf{h}_i}$ を結合して，ベクトル $\mathbf{h}_i$ を得る:

$$\mathbf{h}_i = [\overrightarrow{\mathbf{h}_i}; \overleftarrow{\mathbf{h}_i}]. \quad (3)$$

合計 u 個の縦ベクトル $\mathbf{h}_i^T$ を横方向に並べて，行列 $\mathbf{H}$ を得る:

$$\mathbf{H} = [\mathbf{h}_1^T \ \cdots \ \mathbf{h}_u^T]. \quad (4)$$

直前の発話テキスト T' についても別の双方向 LSTM で同様に読み込み，行列 $\mathbf{H}'$ を得る．LSTM の隠れ状態の行列 $\mathbf{H}$ は，次に深さ r の自己注意機構 [17] に渡される:

$$\mathbf{A} = \mathtt{Softmax}(\mathbf{W}_2 \mathtt{tanh}(\mathbf{W}_1 \mathbf{H})), \quad (5)$$

$$\mathbf{M} = \mathbf{A}\mathbf{H}^T. \quad (6)$$

ここで，$\mathbf{A}$ は自己注意機構の重み行列であり，行列 $\mathbf{W}_1, \mathbf{W}_2$ を用いて推定される．$\mathbf{W}_1, \mathbf{W}_2$ のサイズは，LSTM の隠れ

状態の次元数を d_{lstm}，自己注意機構の深さを r として，それぞれ $d_a \times 2d_{lstm}, r \times d_a$ となる．ここで，d_a は自己注意機構の重みを推定するニューラルネットワークの行列のサイズを制御するパラメータである．自己注意機構の出力である行列 $\mathbf{M}$ のサイズは $r \times 2d_{lstm}$ となる．行列 $\mathbf{M}$ の各行要素は以下のように結合ベクトル $\mathbf{t}$ に変換される:

$$\mathbf{t} = [\mathbf{m}_1; ...; \mathbf{m}_r]. \tag{7}$$

ここで，$\mathbf{m}_i$ は行列 $\mathbf{M}$ の第 i 行目成分を表す．このようにして得られた $\mathbf{t}$ を対象発話テキスト T に対する最終的なエンコーダ出力とする．直前の発話テキスト T' に対しても同様に別の自己注意機構で読み込み，エンコーダ出力 $\mathbf{t}'$ を得る．

補助的なデータ $D_1, ..., D_3$ については，それぞれを one-hot 表現に変換したものを $\mathbf{d}_1, ..., \mathbf{d}_3$ として，それらを線形変換し埋め込み表現に変換する:

$$\mathbf{d}'_n = \mathbf{d}_n \mathbf{W}_n + \mathbf{b}_n. \tag{8}$$

そして，$\mathbf{t}, \mathbf{t}', \mathbf{d}'_1, ..., \mathbf{d}'_3$ を結合したのち，この結合ベクトルのサイズがラベル種類数と同一サイズのベクトルとなるよう，適当なサイズの行列 $\mathbf{W}_3$ を用いて線形変換する．最後に，sigmoid 関数によって各次元が確率値になるよう正規化され，スコアを表現するベクトル $\mathbf{p}$ を得る:

$$\mathbf{p} = \mathtt{sigmoid}([\mathbf{t}; \mathbf{t}'; \mathbf{d}'_1; \mathbf{d}'_2; \mathbf{d}'_3]\mathbf{W}_3). \tag{9}$$

なお，未来発話ラベル予測タスクにおいては，上式から $\mathbf{t}$ が除去されスコアを表現するベクトル $\mathbf{p}$ を得る．スコアが閾値 t 以上のラベルを最終的に出力する．t は開発セットでの F 値が最も高くなるよう調整する．また，上記の予測モデルは二値交差エントロピー損失を最小化するよう学習される．

4. 実験

実験に用いたデータ，実験設定，ベースライン手法，結果について述べる．

4.1 データ及びパラメータの設定

実験では，テキストデータと補助的なデータの 619 組に対して 5 分割交差検証を行った．訓練データとして 80%，評価データとして 20%を用いる．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数は 300 とした．MeCab の辞書には IPADic を用いた．LSTM の隠れ状態の次元数 d_{lstm} 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 の推定に用いられる行列 $\mathbf{W}_1, \mathbf{W}_2$ のパラメータ d_a 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，その初期値としてランダムに生成したベクトルを用いた．また，予測モデルを学習する際に，単語埋め込み表現のパラメータも更新するように設定した．ネットワークの最適化は Adam [6] の初期学習率を 0.001 に設定し行った．学習時のバッチサイズは 64 とした．視点情報，時間情報，直前ラベル，それぞれの埋め込み層の行列サイズは，$2 \times 2, 4 \times 2, 24 \times 4$ 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．実験は各設定において，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．

4.2 ベースライン手法

本実験の対象とする問題はマルチラベル設定であるため，各発話に対し最頻出ラベル上位 k 個を予測するベースラインを考える．本実験では，各発話に対する k の値の決定方法に関して，2 種類の手法を用いる．1 つ目は k の値をデータセット中の発話に付与された平均ラベル数とする手法である．データセットにおいて，各発話に対して付与されているラベル数の平均値は 1.44 であった．そこで，k の値を 1，あるいは 2 とし，全評価セット内の発話に対して上位 k 番目までの最頻出ラベルを予測する．2 つ目は k を各発話に対して付与されている正解ラベルの数とする手法である．すなわち正解ラベルと同数の上位の最頻出ラベルを予測する．

4.3 発話ラベル予測タスクでのモデルの評価

発話テキストに対するマルチラベル分類問題として，正解ラベルに対する予測ラベルの適合率，再現率，F 値を評価指標として用いる．比較手法の性能を表 3 に示す．対象発話テキストのみを用いるモデル，対象発話テキストと時間，視点などの補助的なデータを用いるモデル，そして，対象発話テキストと補助的なデータに加え，さらに直前のテキストも用いるモデルの 3 つの異なるモデルを比較する．

その結果，ニューラルネットワークによる 3 つのモデルはいずれも F 値において .72 を超えており，ベースラインで最も F 値の高い .360 よりも大幅に良い性能を示した．F 値に関して，対象発話テキストに加えて補助的なデータを用いるモデル，及びさらに直前の発話テキストも用いるモデルは，対象発話テキストのみを用いるモデルに対して性能が低下していることが観察された．この結果より，補助的なデータあるいは直前の発話テキストを用いて，発話テキストのラベル推定の精度向上を試みることには未だ課題があると思われる．詳細な議論は，5 節で行う．

	適合率	再現率	F 値
ベースライン:			
k-最頻出ラベル ($k = 1$)	.351	.243	.287
k-最頻出ラベル ($k = 2$)	.297	.412	.345
k-最頻出ラベル (k は正解ラベル数)	.360	.360	.360
対象発話ラベル予測タスク:			
テキストのみ	.773±.024	.706±.023	.735±.004
テキスト + 視点 + 時間 + 直前ラベル	.769±.013	.694±.017	.727±.008
テキスト + 視点 + 時間 + 直前ラベル + 直前テキスト	.762±.031	.691±.025	.722±.010
未来発話ラベル予測タスク:			
視点 + 時間 + 直前ラベル + 直前テキスト	.213±.005	.514±.021	.298±.004

表 3 対象発話ラベル予測タスク及び未来発話ラベル予測タスクでの性能評価.

4.4 未来発話ラベル予測タスクでのモデルの評価

補助的なデータ及び直前の発話テキストを用いるモデルを評価する．対象発話テキストを含む設定と同様に，正解ラベルと推定ラベル間の適合率，再現率，F 値を評価し，ベースラインとして最頻出ラベルを使用する手法を用いる．適合率，再現率，F 値を表 3 に示す．F 値に着目すると，ベースラインの最高性能は .360 であるのに対し，ニューラルネットワークを用いた手法は .298 であり，ベースラインよりも低い性能を示した．

5. 議論

対象発話ラベル予測タスクにおいて，対象発話テキストに加えて，補助的なデータや直前の発話テキストを用いる手法では，直前の発話テキストを用いない手法と比べて，性能が低下することが確認された．これは，本実験の発話分類タスクでは，分類対象発話のテキストが非常に大きな手掛かりとなっており，他の情報はノイズとなっている可能性が考えられる．

未来発話ラベル予測タスクにおいて，本稿で提案された手法はベースライン手法よりも低い性能を示した．多くの発話ラベルは対象発話テキストの情報に大きく依存しており，対象発話テキストを与えない未来発話ラベル予測タスクにおいては，シンプルな手法では k-最頻出ラベルのベースラインよりも性能が低く，非常に難しいタスクであることが分かった．

一方で，レーシングゲーム実況において利用可能である，映像データや，車やサーキットの構造化データは本稿では対象としておらず，これらを組み合わせるマルチモーダルな手法の開発は主となる今後の研究の方向性である．また，次の時点における発話内容は人間の実況者においてもしばしば異なることがあり，未来発話ラベル予測タスクにおけるモデルの性能の上限を探るために，人間の実況者間における次の発話ラベルの一致率がどの程度かを調査する必要があると考えられる．

6. おわりに

本実験では，実況テキストにおける発話ラベルの予測を二通りの設定で行った。分析より，対象発話ラベル予測タスクにおいては，ニューラルネットワークを用いた手法は単純なベースラインよりも良い性能を示すものの，発話の視点情報や時間情報，直前の文脈情報を用いて，ラベル予測の精度向上を計ることは挑戦的であることが分かった．また，発話テキストを用いない未来発話ラベル予測タスクにおいては，提案手法はベースライン手法と比較して劣化した性能を示した．ここからも，テキストデータに加えて視点情報や時間情報などの補助的なデータを用いるマルチモーダルな手法は挑戦的であることが分かった．また，未来発話ラベル予測タスクにおけるモデルの性能の上限を探るために，次の時点での発話ラベルに対する人間の実況者間の一致率を調べる必要があることを述べた．

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] Austin, J. L.: *How to do things with words*, William James Lectures, Oxford University Press (1962).
[2] Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, *Neural Computation*, Vol. 9, No. 8, pp. 1735–1780 (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)*, pp. 103–113 (2021).
[4] Khanpour, H., Guntakandla, N. and Nielsen, R.: Dialogue act classification in domain-independent conversations using a deep recurrent neural network, *Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers (COLING2016)*,

pp. 2012–2021 (2016).
[5] Kim, B. J. and Choi, Y.: Automatic baseball commentary generation using deep learning, *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, p. 1056–1065 (2020).
[6] Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization (2014).
[7] 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, pp. 527–534 (2013).
[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 (1983).
[9] Liang, P. P., Lim, Y. C., Tsai, Y.-H. H., Salakhutdinov, R. and Morency, L.-P.: Strong and Simple Baselines for Multimodal Utterance Embeddings, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL2019)*, pp. 2599–2609 (2019).
[10] 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 (SIGDIAL2014)*, pp. 133–140 (2014).
[11] 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 (2019).
[12] Searle, J. R.: *Speech Acts: An Essay in the Philosophy of Language*, Cambridge University Press (1969).
[13] 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 (2000).
[14] 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)* (1998).
[15] 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 (2019).
[16] 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)*, pp. 7125–7134 (2020).
[17] Zhouhan, L., Minwei, F., dos Santos Cicero Nogueira, Mo, Y., Bing, X., Bowen, Z. and Yoshua, B.: A Structured Self-attentive Sentence Embedding, *Proceedings of the International Conference on Learning Representations (ICLR)* (2017).