

## 数値気象予報からの天気予報コメントの自動生成

村上聡一郎<sup>†</sup>・田中 天<sup>†</sup>・萩行 正嗣<sup>††</sup>・上垣外英剛<sup>†</sup>・  
船越孝太郎<sup>†</sup>・高村 大也<sup>††</sup>・奥村 学<sup>†</sup>

本研究では、数値気象予報のシミュレーション結果から天気予報コメントを自動生成するタスクに取り組む。天気予報コメントの生成タスクには、(i) 様々な物理量の数値変化を考慮する必要がある、(ii) コメントの配信時刻や対象エリアに依存した表現が使われる、(iii) 天気予報コメントにおいて情報の有用性が重要視されている、といった特徴的な課題がある。本研究では、数値気象予報のシミュレーション結果、気象観測値、コメントのメタ情報を入力として、上記の特徴を捉えた上でテキスト化するための Data-to-Text モデルを提案する。また、天気予報コメントにおける情報の有用性の向上のために、晴天や雨などの気象情報を表す「天気ラベル」を予測する内容選択モデルを導入し、予測結果をテキスト生成時に考慮することで有用な情報を明示的に記述できるようにした。実験では、自動評価と人手評価を行い、提案モデルはベースラインに対して情報の有用性の観点で最も優れていることを示した。

キーワード：Data-to-Text, 自然言語生成, ニューラルネットワーク

## Generating Weather Comments from Numerical Weather Prediction

SOICHIRO MURAKAMI<sup>†</sup>, SORA TANAKA<sup>†</sup>, MASATSUGU HANGYO<sup>††</sup>, HIDETAKA KAMIGAITO<sup>†</sup>, KOTARO FUNAKOSHI<sup>†</sup>, HIROYA TAKAMURA<sup>†,††</sup> and MANABU OKUMURA<sup>†</sup>

This research investigates the task of generating weather-forecast comments from simulation results of numerical weather prediction. This task has the following requirements. (i) The changes in numerical values for various physical quantities must be considered; (ii) the weather comments should be dependent on delivery time and area information; and (iii) the comments should provide useful information for users. To meet these requirements, we propose a data-to-text model, incorporating three types of encoders for numerical forecast maps, observation data, and metadata. We also introduce *weather labels* representing weather information, such as sunny or rain, in our model to describe useful information explicitly. Furthermore, we conducted automatic and human evaluations. The results indicate that our model exhibits the best performance when compared with baseline models in terms of informativeness.

---

<sup>†</sup> 東京工業大学, Tokyo Institute of Technology

<sup>††</sup> 産業技術総合研究所, National Institute of Advanced Industrial Science and Technology

<sup>†††</sup> 株式会社ウェザーニューズ, Weathernews Inc.

**Key Words:** *Data-to-Text, Natural Language Generation, Neural Networks*

## 1 はじめに

近年の天気予報は, ある時点の気象観測データと大気の状態に基づいて, 風や気温などの時間変化を数値モデルによりコンピュータで計算し, 将来の大気の状態を予測する数値気象予報 (Numerical Weather Prediction; NWP) が主流となっている. ウェザーニュース<sup>1</sup>や Yahoo!天気<sup>2</sup>の天気予報サイトでは, 数値気象予報に基づき作成された天気図や表データと共に, 気象情報をユーザーに分かりやすく伝えるための天気予報コメントが配信されている. これらの天気予報コメントは, 数値気象予報や過去の気象観測データ, 専門知識に基づいて気象の専門家により記述されている. また, 天気予報サイトでは, 特定のエリアや施設周辺に限定して天気予報を伝えるピンポイント天気予報が一般的になっている. 一方で, 全国の天気予報コメントを作成するのは手間がかかる上に, 専門的な知識を要するため作業コストが高い. そのため, 自然言語生成の分野では, 天気予報コメントの自動生成タスクについて長年取り組まれている (Goldberg et al. 1994; Belz 2007).

本論文では, 数値気象予報のシミュレーション結果から天気予報コメントを生成するタスクに取り組む. これまで取り組まれてきた天気予報コメント生成の研究では, 数値気象予報のシミュレーション結果から気象の専門家の知識と経験に基づき作成した構造化データを用いた研究が中心であったが (Reiter et al. 2005; Sripada et al. 2004; Liang et al. 2009), 本研究では, 数値気象予報の生のシミュレーション結果を用いる. これは, 気象の専門家が数値気象予報から天気予報コメントを記述する実際のシナリオに近い設定であり, 天気予報コメントの作成作業の自動化においても有用であると考えられる.

ここで, 図1を用いて, 天気予報コメントの生成における特徴的な3つの問題について説明する.

まず, 第一の問題は, コメントを記述する際に降水量や海面更正気圧等の複数の物理量とそれぞれの時間変化を考慮しなければならないことである. 例えば, 図1では, 降水量や雲量といった複数の物理量の時間変化に応じて, 日差しが出た後に雲が広がり雨が降ることについて言及されている. 次に, 第二の問題は, 天気予報コメントは, 対象となる地域やコメントの配信時刻, 日付といったメタ情報に基づいて記述されることである. 例えば, 午前中に配信される天気予報コメントでは, 図1のように, 配信日当日の日中から夕方にかけての天気に関及することが多く, 夕方以降に配信される天気予報コメントでは, 配信日当日の夜から翌日の日中の天気に関及する傾向がある. 最後に, 第三の問題は, 天気予報サイトのユーザーは天気予報コメントの情報の有用性 (以降では, 情報性と呼称する) を重要視している点である. 特に, 「晴れ」「雨」「曇り」「雪」といった気象情

<sup>1</sup> <https://weathernews.jp/>

<sup>2</sup> <https://weather.yahoo.co.jp/weather/>

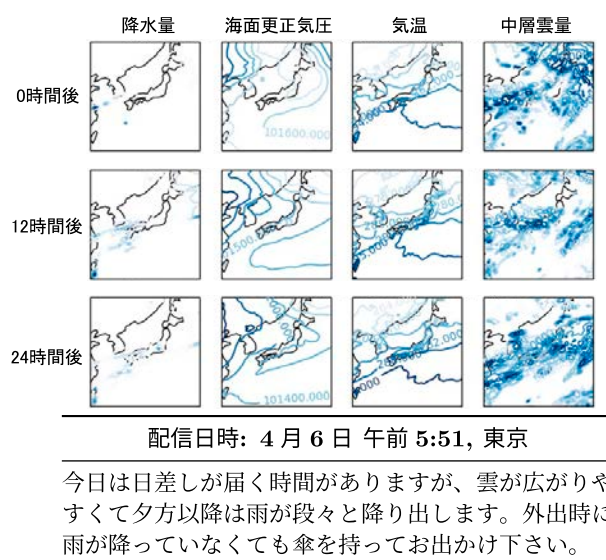


図 1 数値気象予報のシミュレーション結果と天気予報コメントの例

報は、ユーザーの服装や予定に大きな影響を与えることから明示的に記載する必要がある。例えば、図 1 では、降水量、雲量、気温など、記述すべき内容はいくつか考えられるが、雨や傘の情報はユーザーの行動に大きな影響を与えるため、主に雨や傘の情報に焦点を当てている。

これらの問題に対して、本研究では数値気象予報のシミュレーション結果から天気予報コメントを生成するための Data-to-Text モデルを提案する。第一の問題に対しては、多層パーセプトロン (Multi-Layer Perceptron; MLP) や畳み込みニューラルネットワーク (Convolutional Neural Network; CNN) を用いて様々な物理量を捉え、それらの時間変化を双方向リカレントニューラルネットワーク (Bidirectional Recurrent Neural Network; Bi-RNN) を用いて考慮する。第二の問題については、エリア情報やコメントの配信時刻、日付などのメタ情報を生成モデルへ取り入れることでこれらの情報を考慮する。第三の問題について、本研究では「晴れ」「雨」「曇り」「雪」に関する気象情報をユーザーにとって重要な情報と定義し、これらを適切に言及するための機構を提案する。具体的には、言及すべき重要な情報を明示的に記述するために、数値気象予報のシミュレーション結果から「晴れ」「雨」「曇り」「雪」の気象情報を表す「天気ラベル」を予測する内容選択モデルを導入し、予測結果をテキスト生成時に考慮することで、生成テキストの情報性の向上に取り組む。

実験では、数値気象予報のシミュレーション結果、気象観測データ、および、人手で書かれた天気予報コメントを用いて提案手法の評価を行った。自動評価では、人手で書かれた天気予報コメントと生成テキストの単語の一致度合いを評価するための BLEU および ROUGE、また、生成テキス

トにおいて天気ラベルが正確に反映されているかを評価するための F 値を使用し, 提案手法がベースライン手法に比べて性能が改善することを確認した. さらに, 人手評価では, 提案手法はベースライン手法と比較して, 天気予報コメントの情報性が向上していることが示された.

## 2 関連研究

時系列数値データや構造化データといった非言語データを人間にとって分かりやすく説明するために, 様々なドメインにおいて Data-to-Text タスクの研究が取り組まれている. 例えば, バスケットボールの試合における選手のスタッツやボックススコアからなる表データから試合要約テキストを生成する研究 (Wiseman et al. 2017; Puduppully et al. 2019), 日経平均株価からなる時系列数値データから株価の概況テキストを生成する研究 (Murakami et al. 2017; Aoki et al. 2018, 2019), レストランの概要を表すメタ情報から説明テキストを生成する研究 (Novikova et al. 2017) などが行われている. 本研究で取り組む気象シミュレーションからの天気予報コメントの生成タスクは, Data-to-Text タスク (Gatt and Krahmer 2018) の一種である.

一般的に Data-to-Text タスクは, (i) 入力データの中で言及すべき内容を選択する **内容選択** (*content selection*), (ii) 選択した内容をどのように言及するかを表す内容プラン (*content plan*) を作成するための **内容プランニング** (*content planning*), (iii) 実際のテキストを生成する **表層化** (*surface realization*) の 3 つのサブタスクにより構成される (Gatt and Krahmer 2018). 従来, これらのサブタスクは, 知識や経験に基づくルールベース (Kukich 1983; Reiter and Dale 1997; Reiter et al. 2005) や過去のデータから各サブタスクの規則を獲得する機械学習ベース (Duboue and McKeown 2001, 2003; Barzilay and Lapata 2005) などの手法により, それぞれ個々に取り組まれてきた. 一方, 近年では, 情報通信分野の発展によりウェブから大規模データの収集が容易になったことで, 入力データと出力テキストのペアから機械学習によりテキスト生成規則を自動的に獲得する End-to-End な手法が広く用いられている (Liu et al. 2018; Iso et al. 2019). 特に最近では, 機械翻訳分野で提案されたニューラルネットワークに基づくエンコーダ・デコーダモデル (Bahdanau et al. 2015; Sutskever et al. 2014; Cho et al. 2014) を用いた手法に注目が集まっており, 様々な Data-to-Text タスクにおいて有用性が報告されている (Sha et al. 2018; Mei et al. 2016; Lebrete et al. 2016). 本研究では, 近年取り組まれているこれらの研究と同様に, 天気予報コメント生成タスクを数値気象予報のシミュレーション結果から天気予報コメントを生成する系列生成タスクとして考え, エンコーダ・デコーダを用いた手法を提案する.

天気予報コメントは, 一般的なユーザー向けの天気予報コメントと専門的な天気予報コメントの 2 種類に大別できる. 一般的なユーザー向けの天気予報コメントとは, ウェザーニュースや Yahoo! 天気といった天気予報サイトで配信されている一般ユーザー向けの天気予報コメントのことを表す. また, 専門的な天気予報コメントとは, 海運や農業, 航空業界といった特定の業界向けの専門的な

天気予報コメントのことである。そのため、天気予報コメントは、対象とするユーザーや業種によって書かれる内容は様々である。例えば、一般的なユーザー向けの地域の天気予報コメントの生成の研究 (Kerpedjiev 1992; Liang et al. 2009) では、雲の量や雨の時間帯、風の強さ、気温といった幅広い観点について言及する天気予報コメントの生成を対象としている。一方、海運や海洋石油施設を対象とした海上の天気予報コメントの生成の研究 (Kittredge et al. 1986; Reiter et al. 2005) では、海上の風の強さや波の高さを中心に言及するコメントの生成を対象としている。本研究では、ウェザーニュースや Yahoo!天気などの天気予報サイトで配信されている一般的なユーザー向けの天気予報コメントの生成を対象としている。

天気予報コメントの自動生成タスクは、Data-to-Text の分野において長年取り組まれている課題の 1 つである (Belz 2007; Angeli et al. 2010; Mei et al. 2016)。これまでの研究において、天気予報コメント生成タスクを対象としたデータセットである、図 2 の SUMTIME-METEO (Sripada et al. 2003) や図 3 の WEATHERGOV (Liang et al. 2009) といったデータセットが公開され、様々な研究で広く用いられている。これらのデータセットは、数値気象予報のシミュレーション結果を専門家の知識・経験に基づき修正した表やデータベース形式の構造化データと天気予報コメントのテキストデータから構成されている。SUMTIME-METEO は、SUMTIME と呼ばれる時系列データの概況テキストの生成技術に関する研究プロジェクト (Sripada et al. 2002) において作成された

#### 気象データ:

Time	Wind dir	Wind speed	Gust 10m	Gust 50m	Sig. Wave Height	Wave Period
01:00	SE	21	26	32	2.10	3.40
04:00	ESE	17	21	26	2.10	3.40
07:00	E	18	22	28	2.00	3.20
10:00	E	16	20	24	1.90	3.00
13:00	ENE	16	20	24	1.90	3.00
16:00	NE	14	17	21	1.90	3.00
19:00	NE	16	20	24	1.50	2.40
22:00	NNE	16	20	24	1.50	2.40

#### 天気予報コメント:

Field	Text
Wind at 10m	E-SE 18-22 GRADUALLY BACKING/EASING NNE 15-20
Wind at 50m	E-SE 22-28 GRADUALLY BACKING/EASING NNE 18-25
Sig. Wave Height	AROUND 2.0 GRADUALLY FALLING 1.5-2.0
Max Wave Height	AROUND 3.0 GRADUALLY FALLING 2.5-3.0

図 2 SUMTIME-METEO の例

気象データ:

```
skyCover(date=2009-02-07, label=Tonight, time=17-30, mode=50-75)
temperature(date=2009-02-07, label=Tonight, time=17-30, min=-6, mean=-1, max=8)
windDir(date=2009-02-07, label=Tonight, time=17-30, mode=S)
windSpeed(date=2009-02-07, label=Tonight, time=17-30, min=2, mean=3, max=5, mode=0-10)
```

天気予報コメント:

*“Mostly cloudy, with a low around -6. South wind around 5 mph becoming calm.”*

図 3 WEATHERGOV の例

北海における海洋石油施設向けの海洋気象を対象としたデータセットである<sup>3</sup>。本データの特徴として、図 2 のように、海上の風や波の高さに関する時系列データと風や波などのそれぞれの物理量に対する人手で書かれた短文テキストから構成されていることが挙げられる。また、WEATHERGOV は、米国の都市を対象とした天気予報配信サイト Weather.gov<sup>4</sup>から収集されたデータセットである。本データの特徴としては、図 3 のように、気温や風向き、雲量等に関するデータベース形式の構造化データとそれらの物理量全体について概況する天気予報コメントから構成されていることが挙げられる。また、WEATHERGOV の天気予報コメントの多くは、ルールベースに基づくシステムにより自動生成されたテキストやそれらを人手で修正したテキストであることから、一般的な天気予報サイトにおける天気予報コメントと比べて単調な文章であることが知られている (Wiseman et al. 2017)。

本研究では、これまでの天気予報コメント生成に関する研究で扱われてきた SUMTIME-METEO や WEATHERGOV といった構造化データではなく、数値気象予報のシミュレーション結果、および、雨量、日照時間等の気象観測値といった時系列数値データを入力として考える。これは、気象の専門家が天気予報コメントを記述する際にこれらの時系列数値データを参照するという実際のシナリオを模したタスク設定となっている。このように、数値気象予報のシミュレーション結果や気象観測値等の生データを入力として考えることで、天気予報コメントの自動生成タスクにおいて次の 2 つの利点があると考えられる。まず、1 つ目に、気象の専門家が参照する数値気象予報のシミュレーション結果や気象観測値等の生データは情報量が多く、SUMTIME-METEO や WEATHERGOV のようにデータを構造化することで情報量が落ちることが懸念されるが、生データを用いることで、含まれる情報を全て活用することが可能となる。2 つ目に、数値気象予報のシミュレーション結果から人手またはシステムを介して事前に構造化データを作成するという手間がないため、天気予報

<sup>3</sup> 当該研究プロジェクトにおけるその他のデータセットとして、ガスタービンや新生児集中治療室におけるセンサー値とそれらの概況テキストから構成される SUMTIME-TURBINE と SUMTIME-NEONATE がある。

<sup>4</sup> <https://www.weather.gov/>

コメントの作成作業の自動化やシステム構成の簡易化の観点から有用である。これは、我々がテレビや新聞で目にする一般的な天気予報の情報（例えば、曇り時々雨）においても同様である。一般的な天気予報の情報の一部は、気象庁の予報官や気象会社の気象予報士といった専門家が数値気象予報の結果や観測値等に基づいて作成しているため、これらをシステムの入力として前提した場合、人手を介した作業が必要となり、天気予報コメントの作成作業の自動化において障壁になることが懸念される。そのため、本研究では、一般的な天気予報の情報や構造化データを入力として使用していない<sup>5</sup>。

一方で、本研究で扱う気象データは、これまでの研究で扱われてきた構造化データと形式が大きく異なることから、従来研究で提案されてきた構造化データからテキスト生成するための Table-to-Text 技術 (Mei et al. 2016; Wiseman et al. 2017; Ma et al. 2019) を時系列数値データへ直接的に適用することが難しい場合がある。加えて、Data-to-Text 生成の分野全体においても、時系列数値データからのテキスト生成に関する取り組みは限られている (Gatt and Krahmer 2018)。したがって、本タスクの第一の問題として挙げた複数の物理量からなる時系列数値データを捉えてテキスト化するためのエンコード/デコード手法の考案・検証が求められる。

また、Data-to-Text 生成の分野において、本研究で取り組む第二および第三の問題に関連する研究はいくつか取り組まれている (Li et al. 2016; Murakami et al. 2017; Puduppully et al. 2019; Wang et al. 2020; Ma et al. 2019; Lebrete et al. 2016)。例えば、第二の問題として挙げた、テキストの配信時刻や日付、エリア等のメタ情報に基づきテキストが書かれる特徴については、時系列株価データの市況コメントを生成する Data-to-Text 生成課題において、テキストのメタ情報（時間帯情報）を表す埋め込みベクトルをデコーダに組み込むことでメタ情報を捉えた単語列を生成できることが報告されている (Murakami et al. 2017)。この研究は、Li et al. (2016) が行ったエンコーダ・デコーダモデルを用いた対話システムの研究で、デコーダにペルソナ情報を与えることで、指定したペルソナの特徴を捉えた単語列が生成が可能になるという報告に基づいた取り組みとなっている。本研究では、これらの研究に倣い、テキストの配信時刻や日時、エリア等のメタ情報を表す埋め込みベクトルをモデルへ組み込むことで、メタ情報を捉えた天気予報コメントの生成に取り組む。

第三の問題として挙げた、ユーザーにとって天気予報コメントの情報は重要であるという特徴は、配信テキストの信憑性やユーザー満足度などの観点から、本研究で取り組む天気予報コメント生成に限らず Data-to-Text 生成の分野全体において極めて重要な課題である。そのため、Data-to-Text 生成の分野において、入力データの事実に基づく正確なテキスト生成に関する研究が注目されている。特に、近年のエンコーダ・デコーダモデルに基づくテキスト生成モデルでは、内容選択や内容プランニングを明示的に行う外部機構を導入することでテキストの情報性向上に取り組む動きが盛ん

<sup>5</sup> その他の一般的な天気予報の情報として「降水確率」や「最高気温」、「最低気温」などがあるが、入力データとして使用した数値気象予報のシミュレーション結果には「降水量」や「気温」の予測値といった類似した情報が含まれていることから、今回は不要と判断し、使用していない。

になっており, 有用性が報告されている (Puduppully et al. 2019; Wang et al. 2020; Ma et al. 2019; Lebrete et al. 2016). 例えば, Table-to-Text タスクの研究では, 表データに含まれる固有名詞や数詞等を正確に記述するためにコピー機構 (Luong et al. 2015) を導入する手法 (Lebrete et al. 2016) や, 表データの中で言及すべき内容を予測し, 明示的に生成モデルに組み込む手法 (Ma et al. 2019), 表データの中から言及すべきエンティティや数字の情報抽出し, これらを基に作成した内容プランを生成モデルへ導入することで正確かつ一貫性のあるテキストを生成する手法 (Puduppully et al. 2019) などが提案されている. しかしながら, これらの手法では, 表から単語をコピーするためのコピー機構や表から言及すべき内容を予測する分類器や情報抽出システムのように, Table-to-Text タスク特有の外部機構が必要となる. また, これらの分類器や情報抽出システムを学習するためには, 表データとテキストの単語一致に基づいて教師データを事前に作成する必要がある (Wiseman et al. 2017). そのため, 本研究で取り組む時系列数値データからのテキスト生成タスクでは, これらの研究と入力データの性質が大きく異なるため, Table-to-Text タスク特有の外部機構の導入が困難となる. このような状況を踏まえ, 本研究では, 入力データを参照せずにテキスト側から言及すべき内容を抽出することにより, 内容選択を明示的に行う分類器のための教師データを作成する. これにより, 外部機構として内容選択を行う分類器の学習が可能となった. なお, 以降では, 外部機構として内容選択を行う分類器のことを内容選択モデルと呼称する.

### 3 気象データの概要

天気予報コメントは, 気象の専門家が数値気象予報や過去の気象観測データ, 気象の専門知識を基に記述している. これに従い, 本研究では数値気象予報のシミュレーション結果である数値予報マップと気象観測データを使用した. 本章では, これらの詳細について解説する.

#### 3.1 数値予報マップ

数値予報マップとは, 数値気象予報モデルを大規模コンピュータで三次元シミュレーションした結果から, 地表部分を取り出した二次元面データである. 本研究では, 数値気象予報モデルの一種である全球数値予報モデル (Global Spectral Model; GSM) を用いて計算された日本周辺の数値予報マップを使用する. 表 1 に本研究で使用する日本周辺の数値予報マップの概要を示す. 数値予報マップは, 気圧や気温, 風向きなどの各物理量の 1 時間ごとの予測数値が 84 時間先まで格納された時系列数値データである. 気象庁が作成している日本周辺の数値予報マップには, 北緯 20 度から 50 度, 東経 120 度から 150 度の範囲で 20km ごとに格子点が設定されており, 合計 151×121 の格子点から構成されている. また, 各格子点には各物理量の予測数値が格納されている. 例えば, 気圧の場合は 1021.01hPa, 気温の場合は 258.52K 等の数値が含まれている.

図 4 に降水量, 海面更正気圧, 気温, 中層雲量を日本周辺の地図上に可視化した数値予報マップ



配信時刻	6 時間毎に更新 (1 日 4 回: 00, 06, 12, 18UTC)
予報時間	84 時間予報
領域	北緯 20 度から 50 度, 東経 120 度から 150 度
物理量 (11 種類)	気圧, 海面更正気圧, 東西風, 南北風, 気温, 相対湿度, 降水量, 上層雲量, 中層雲量, 低層雲量, 全雲量

表 1 数値予報マップの概要

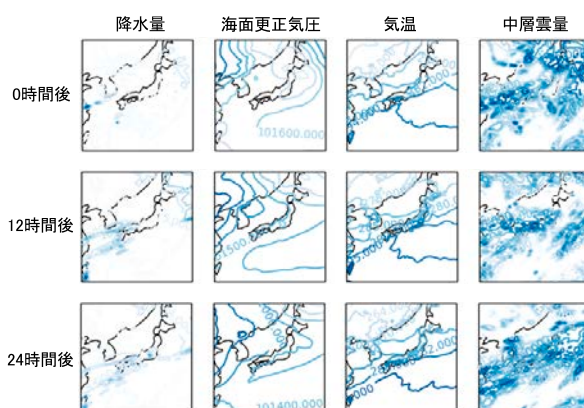


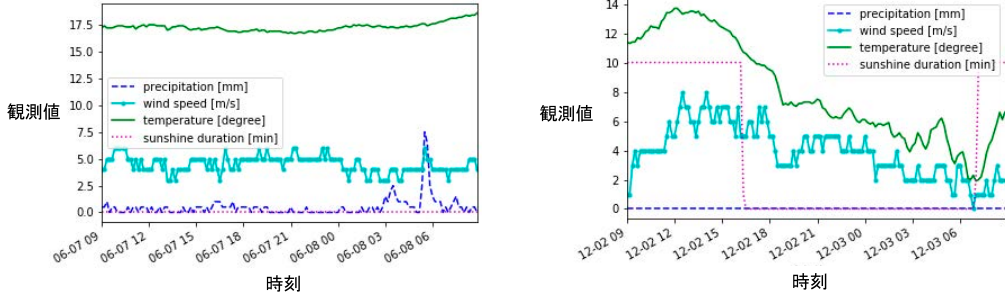
図 4 数値予報マップの例

の例を示す. 今回使用した数値予報マップでは, 数値気象予報のシミュレーションの起点となる時刻の 0 時間後 (直後) から 84 時間後までの各物理量の予測数値が含まれているが, 図 1 では 0 時間後から 12 時間後, 24 時間後の予測数値を可視化している. また, 可視化した数値予報マップ上の色の濃淡は予測数値の大きさを表している. 例えば, 中層雲量の色が濃い箇所は雲量が多いことを表し, 色が薄い箇所は雲量が少ないことを表している.

### 3.2 気象観測データ

気象観測データとして, 気象庁の地域気象観測システム<sup>6</sup>(automated meteorological data acquisition system; AMeDAS) から収集された観測値を使用した. AMeDAS は, 全国の約 1,300 地点に設置されており, 10 分毎の降水量, 気温, 風, 日照時間を計測している. 図 5 に「所沢」エリアの AMeDAS による 10 分毎の観測値データの例を示す. 図 5(a) および図 5(b) はそれぞれ 2014 年 6 月 7 日, 2014 年 12 月 2 日の午前 9 時から翌午前 8 時 50 分に観測された 10 分毎の降水量 (precipitation), 気温 (temperature), 風 (wind speed), 日照時間 (sunshine duration) である. こ

<sup>6</sup> <https://www.jma.go.jp/jp/amedas/>



(a) 2014 年 6 月 7 日午前 9 時から翌午前 8 時 50 分 (b) 2014 年 12 月 2 日午前 9 時から翌午前 8 時 50 分

図 5 「所沢」エリアにおける 10 分毎の観測値の例

ここで、降水量は 10 分間における雨量 (mm)、日照時間は 10 分間における日照時間 (分) を表している。

#### 4 提案手法

近年、Data-to-Text タスク (Mei et al. 2016; Lebrete et al. 2016) や動画キャプション生成 (Yao et al. 2015; Long et al. 2018) などのさまざまな系列生成タスクにおいて、機械翻訳分野で注目されているニューラルネットワークに基づくエンコーダ・デコーダモデル (Sutskever et al. 2014; Cho et al. 2014) を用いた研究が提案され、有用性が示されている。本研究では、天気予報コメントの生成を、数値予報マップからなる時系列データから単語系列を生成する系列生成タスクとして考え、注意機構付きのエンコーダ・デコーダモデル (Bahdanau et al. 2015) を用いた手法を提案する。

図 6 に提案モデルの概要を示す。提案モデルでは、数値予報マップからなる時系列データ  $\mathbf{g} = (g_i)_{i=1}^{|\mathbf{g}|}$ 、降水量や気温などの過去の観測データ  $\mathbf{a} = \{a_i\}_{i=1}^{|\mathbf{a}|}$ 、配信日時や対象エリア等のコメントに関するメタ情報  $\mathbf{m} = \{m_i\}_{i=1}^{|\mathbf{m}|}$  の 3 種類のデータを入力とし、天気予報コメント  $\mathbf{w} = (w_i)_{i=1}^{|\mathbf{w}|}$  および言及すべき重要な情報を表す天気ラベル  $\mathbf{l} = \{l_i\}_{i=1}^{|\mathbf{l}|}$  を出力とする。ここで、入力データの  $g_i$ 、 $a_i$ 、 $m_i$  は、数値予報マップ、降水量や日照時間等の観測データを表す数値ベクトル、エリア名や配信日時等のメタ情報を表す埋め込みベクトルをそれぞれ表す。また、出力データの  $w_i$  および  $l_i$  は、生成テキストにおける単語および天気ラベルを表している。

数値予報マップからなる時系列データのエンコーダとして、数値予報マップに含まれる気圧や雲量といった様々な物理量から特徴を抽出するために MLP また CNN を使用し、Bi-RNN により時系列情報を考慮する。気象観測データおよびメタ情報のエンコーダには、MLP を使用する。また、入力データから言及すべき重要な情報を選定する内容選択モデルとして、エンコーダの出力状態ベ

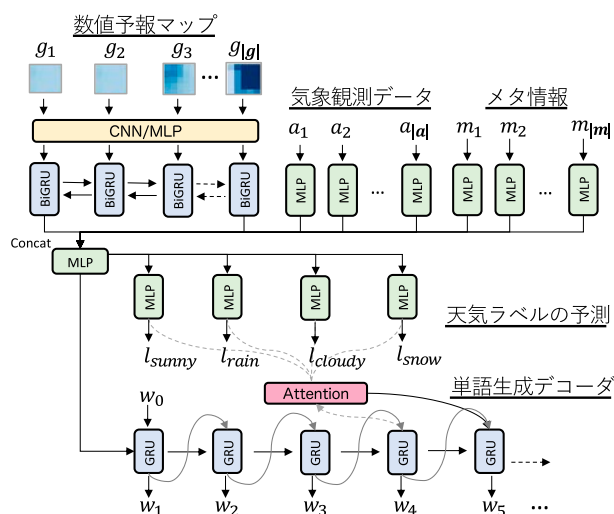


図 6 提案モデルの概要

クトルを入力とする MLP を使用する。天気予報コメントのための単語生成デコーダとして、機械翻訳や文書要約等の系列生成タスクで広く用いられている RNNLM (Mikolov et al. 2010) を使用する。以降では、提案モデルの詳細について説明する。

## 4.1 エリアごとの数値予報マップの抽出

天気予報コメントの作成において、コメント作成者は日本周辺全体の数値予報マップを参照するが、各エリアごとのコメントを記述する際には、主に対象エリア周辺の気象情報に着目して記述することが一般的である。そこで本研究では、各エリアの天気予報コメントは対象エリア周辺の数値気象予報に深く関係しているという仮定のもと、 $151 \times 121$  の格子点からなる日本周辺全体の数値予報マップから、緯度・経度の位置情報を基に対象エリアを中心とする  $5 \times 5$  のマップを抽出し、各エリアの気象情報として利用する。抽出した各エリアの数値予報マップは、対象エリアを中心とした 10,000 平方 km のマップとなる。

ここで、図 7 に東京を対象エリアとし、日本全体の数値予報マップからエリア周辺の数値予報マップを抽出した例を示す。図 7 の例は、降水量 (Prec.), 海面更正気圧 (Pres.), 気温 (Temp.), 総雲量 (Clo), 上層雲量 (h-Clo), 低層雲量 (l-Clo), 中層雲量 (m-Clo), 相対湿度 (Hum.), 東西風 (x-win.), 南北風 (y-win.) の 10 種類の物理量について、24 時間先まで 3 時間ごとの計 9 タイムステップからなる数値予報マップを表している。数値予報マップに含まれる各物理量の予測値は、1 年間の各物理量の予測値の平均および標準偏差を用いて標準化している。また、各物理量の色の濃淡は、予測値の大きさを表している。すなわち、図 7 の 21–24 時間後 (タイムステップ 7 から 8) の降水量 (Prec.) の色の濃さは、降水量の予測値が高いことを示している。

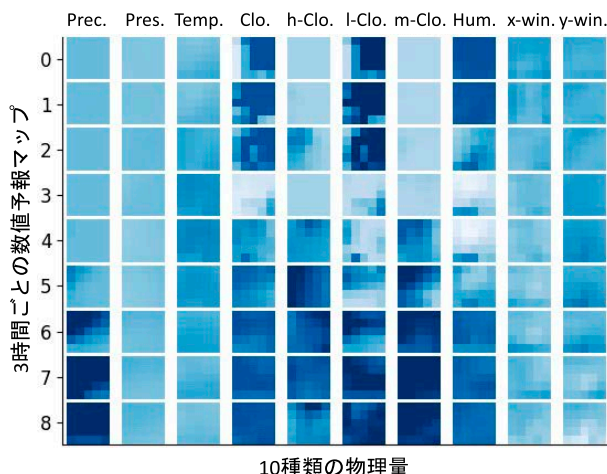


図 7 日本全体の数値予報マップから抽出した東京エリア周辺の数値予報マップ

## 4.2 数値予報マップのエンコード手法

本研究では、天気予報コメント生成タスクを、図 7 に示す時系列の二次元面データである数値予報マップから単語系列を生成する系列生成タスクとして考える。これは、時系列の二次元画像データからなる動画の説明テキスト（キャプション）を生成する動画キャプション生成タスクと類似したタスクとして考えることができる。そこで、本研究では、動画キャプション生成タスクにおいて一般的な CNN または MLP を数値予報マップのエンコーダとして採用し、それらの有用性を比較検証する。

### 4.2.1 CNN を用いたエンコード手法

画像認識や動画キャプション生成タスクでは、入力動画画像の特徴を抽出する手法として CNN が広く使われている。本研究でも同様に、エリア毎の数値予報マップから数値の特徴や物理量間の関係を獲得するために CNN を用いて特徴抽出を行う。また、画像の場合、CNN では画像の RGB 情報から特徴を抽出するために入力チャンネルとして 3 チャンネル (Red, Green, Blue) を使用しているが、本研究の場合は、数値予報マップに含まれる 10 種類の物理量を考慮するために 10 チャンネルを用いて特徴抽出を行う。

### 4.2.2 MLP を用いたエンコード手法

画像認識や動画キャプション生成では、位置不変性の観点から CNN を用いて特徴抽出を行うことが一般的である。しかし、本研究の場合、エリアごとに抽出した数値予報マップを入力としており、着目しているエリアは常にその中心に位置しているため、マップ上の位置をそのまま考慮した

モデルの方が適している可能性が考えられる．そのため本研究では，CNN の代替として MLP を用いた特徴ベクトルの抽出方法についても検証する．具体的には，10 種類の物理量ごとに  $5 \times 5$  の予測値を入力するために  $10 \times 5 \times 5$  個のユニットの入力層を持つ MLP を用いて特徴抽出を行う．

#### 4.2.3 数値予報マップにおける時系列情報の考慮

複数の物理量からなる数値予報マップの時系列的な変化を捉えて天気予報コメントとしてテキスト化するために，前述の CNN または MLP によりエンコードされた数値予報マップを Bi-RNN へ入力する．具体的には，まず，前述の CNN または MLP を用いたエンコード手法により，各タイムステップ  $i$  の数値予報マップ  $g_i$  から特徴ベクトル  $h_i^g$  を取得する．次に，数値予報マップにおける時系列情報を考慮するために，各タイムステップの特徴ベクトルを Bi-RNN によりエンコードし，それぞれの出力ベクトル  $h_i^g$  を獲得する．最後に，時系列データ全体の変化を捉えるために，Bi-RNN の出力ベクトルの先頭と末尾を下記の式のように結合し， $h^g$  を獲得する：

$$h^g = [h_1^g; h_{|g|}^g], \quad (1)$$

ここで， $[\cdot]$  はベクトルの連結演算子を表す．

#### 4.3 気象観測データの導入

天気予報コメントは，数値気象予報と過去の気象観測データを基に記述されることから，本研究では AMeDAS により収集された気象観測データも入力として使用する．具体的には，まず，降水量や日照時間などの過去 24 時間の時系列の観測値からなる数値ベクトル  $a_i$  をそれぞれ MLP によりエンコードし，特徴ベクトル  $h_i^a$  を取得する．次に，各観測データに関する特徴ベクトルを下記の式のように結合し， $h^a$  を獲得する：

$$h^a = [h_1^a; h_2^a; \dots; h_{|a|}^a]. \quad (2)$$

#### 4.4 メタ情報の導入

気象の専門家が天気予報コメントを記述する際には，コメントが書かれる時間帯の情報やエリア特有の情報を考慮して記述することが一般的である．例えば，午前中または午後配信される天気予報コメントでは，以下の天気予報コメント (a)，(b) のように配信時刻に依存する表現が使われる．

(a) 今日 (土) は日差しが届いても，気温は低空飛行で身体の内まで凍える寒さ。

(b) 明日 は雲の間から時々日差しが届きます。

また，天気予報コメントの対象エリアとして，例えば東京都の場合は，「新宿，世田谷，八王子，町田，お台場，大島，三宅島」などの複数エリアに分けられており，それぞれのエリアの特徴（例

えば、海辺、山間部）が反映されたコメントが配信されている．具体的には、海辺が近いエリアに対しては、以下の天気予報コメント (c) のように、エリアに依存する表現が用いられる．

(c) 波も穏やかでマリンレジャーも楽しめそうです。

そこで本研究では、このような表現を生成するために、配信日時（月、日、曜日、時刻）や対象エリア名（新宿、横浜、石垣島等）といった天気予報コメントのメタ情報を導入する．具体的には、まず、配信日時やエリア名から作成した単語埋め込みベクトルを作成する．次に、それらの埋め込みベクトルを MLP によりエンコードすることで各メタ情報の特徴ベクトル  $h_i^m$  を獲得する．最後に、全てのメタ情報  $\mathbf{m}$  を捉えた出力ベクトル  $h^m$  を下記の式のように作成する：

$$h^m = [h_1^m; h_2^m; \dots; h_{|\mathbf{m}|}^m]. \quad (3)$$

これまで説明した以上が提案モデルの入力データである数値予報マップ、気象観測データ、メタ情報のエンコード手法である．これらの手法で獲得した各特徴ベクトル  $h^g, h^a, h^m$  から、下記の式に基づき内容選択モデルおよび単語生成デコーダの初期状態を設定する：

$$s_0 = \text{ReLU}(\text{MLP}([h^g; h^a; h^m])). \quad (4)$$

#### 4.5 内容選択モデルによる天気ラベルの予測

天気予報コメントでは、天気予報コメントの情報性とその正確さが重要な要素となる．しかし、機械翻訳や文書要約等でも広く使われているニューラルネットワークに基づく生成モデルは、入力データや出力単語系列における長期的な依存関係を捉えることに課題があり、入力データに含まれる重要な情報を失うといった問題があることが知られている．例えば、ニューラルネットワークを用いた機械翻訳モデルでは、原文に含まれている内容が訳文で抜けている訳抜けという課題が広く知られている (Tu et al. 2016)．このような課題は、ニューラルネットワークに基づくテキスト生成システムを実世界へ適用する上での障壁となるため、解決が求められる．このような課題に対して Data-to-Text タスクの分野では、生成テキストの情報性を向上させることを目的に、生成テキストの内容を表す明示的なラベルを導入する手法 (Aoki et al. 2019) や表データに含まれる事実・内容そのものを予測しテキスト生成に導入する手法 (Ma et al. 2019; Puduppully et al. 2019)、入力データと生成テキストに含まれる情報の類似性に関する損失関数を用いる手法 (Wang et al. 2020) などが提案されており、それらの有用性が示されている．

本研究では、これら近年の生成テキストの情報性向上に着目した研究から着想を得て、生成モデルが入力データから言及すべき重要な情報を選定するための外部的な機構を導入し、生成テキストにおける情報性の向上に取り組む．具体的には、Ma et al. (2019) や Puduppully et al. (2019) らの手法のように、入力データにおいて言及すべき内容を予測する内容選択モデルを導入し、それらの

予測結果を明示的に生成モデルへ導入する手法を提案する．これにより，内容選択モデルは生成モデルが重要な情報を言及するための補助的な外部機構として作用し，情報性の高いテキストが生成できるようになることを期待する．しかし，入力データの言及すべき内容を予測する内容選択モデルを構築するためには，入力データの言及すべき内容を表す教師データが必要となる．そのため，Ma et al. (2019) や Puduppully et al. (2019) らが取り組む Table-to-Text タスクの場合，表データと参照テキストにおける表層的な単語の一致に基づいて言及すべき内容を表す教師データを作成している (Wiseman et al. 2017)．一方，本研究で取り組む数値気象予報からの天気予報コメント生成タスクのように，入力データが数値データとなる場合，入力データとテキストの単語一致による手法の適用が困難なため，入力と出力テキストの単語一致に依存しない教師データ作成方法が求められる．そこで本研究では，入力データに依存せずに出力テキスト側から言及すべき内容を抽出することにより，内容選択モデルのための教師データを作成する．

以降では，内容選択モデルの教師データの作成方法，内容選択モデルの学習および推論方法について説明する．

#### 4.5.1 内容選択モデルの教師データの作成

天気予報コメントを参照するユーザーは，主に晴れや雨などの気象情報に高い関心を持っていることがほとんどである．そこで，このような気象情報を天気コメントの「天気ラベル」と定義し，「晴れ」，「雨」，「曇り」，「雪」の 4 種類の天気ラベルを導入する．本研究では，言及すべき重要な情報であるこれらの天気ラベルを内容選択モデルにより予測し，テキスト生成時に天気ラベルの予測結果を考慮することで情報性の向上を試みる．

本研究では，内容選択モデルの教師データを作成するために，表 2 に定義した各天気ラベルの手がかり語に基づいて天気予報コメントから天気ラベルを抽出する<sup>7</sup>．具体的には，天気コメントに手がかり語のいずれかが含まれる場合，その天気予報コメントでは該当する気象情報を言及しているとみなし，天気ラベルを紐付ける．例えば，以下の天気コメント (d) には，手がかり語である「日

天気ラベル	手がかり語
SUNNY	晴れ，日差し，青空，回復，日和，陽気，秋晴れ，晴天，晴れ間，晴れる，太陽，五月晴れ
RAIN	雨，大雨，にわか雨，雷雨，暴風雨，雨風，荒天，台風，傘
CLOUDY	曇り，曇，雲
SNOW	雪，吹雪，小雪，吹雪く，ふぶく

表 2 天気ラベルと対応する手がかり語

<sup>7</sup> これらの手がかり語は，開発データを参考に人手により選定した．

差し」, 「雲」, 「雨」, 「傘」が含まれているため, 「晴れ」, 「曇り」, 「雨」の3つの天気ラベルが紐付けられる。

- (d) 今日は日差しが届く時間がありますが、雲が広がりやすくて夕方以降は雨<sub>下</sub>が段々と降り出します。外出時に雨<sub>上</sub>が降っていなくても傘<sub>上</sub>を持ってお出かけ下さい。

手がかり語に基づいて内容選択モデルの学習データを作成するための本手法は簡易的ではあるものの、本研究で取り組む天気予報コメントの生成タスクにおいて、次の2点の利点があると考えられる。まず、1点目として、数値予報マップのような時系列数値データと天気予報コメントを人手やルール等により明示的に紐付けることは困難であるが、本手法では天気予報コメントのみを参照するため、人手やルールによる天気ラベルを対応付けすることができる。2点目として、数値予報マップにおいて言及すべき内容をアノテーションするためには、気象の専門知識が必要となるが、本手法の場合はテキストのみを参照するため、専門知識がない場合でも比較的容易に重要な情報を判定できるという点である。一方、手がかり語に基づくラベル付けは単語レベルで判定しているため、本来テキストで言及されている意図とは異なるといった正確性に関して懸念が生じる。そこで、ランダムにサンプリングされた100件の天気予報コメントに対して上記の方法で天気ラベルの付与を行い、5人の評価者により正確性を評価した。その結果、全体の96%の天気ラベルが適切であると判定されたことから、ほとんどの事例で正確性に問題がないことが分かった。

#### 4.5.2 内容選択モデルの学習および推論方法

次に、上記の方法で作成した教師データを用いて内容選択モデルを学習する方法について説明する。内容選択モデルは、数値予報マップ、気象観測データ、メタ情報 ( $g$ ,  $a$ ,  $m$ ) をエンコードした結果を受け取り、各天気ラベル ( $l_{sunny}$  等) を予測する二値分類器である。内容選択モデルは、各天気ラベルごとに言及すべきか否かを判定する。内容選択モデルの学習時には、学習データの天気予報コメントから抽出した天気ラベルを用いて、各天気ラベルに対する内容選択モデルを学習する。推論時には、3種類の入力データ ( $g$ ,  $a$ ,  $m$ ) から各天気ラベル ( $l_{sunny}$  等) を予測する。また、生成テキストの情報性の向上を目的として、天気予報コメントの生成時には、テキストで言及すべき内容を表す内容選択モデルの予測結果を単語生成デコーダへ導入する。

#### 4.6 天気予報コメントの生成

天気予報コメントの生成に使用する単語生成デコーダとして、機械翻訳や文書要約等の系列生成タスクで広く用いられているRNNLMを使用する。RNNとして、Gated Recurrent Unit (GRU)(Chung et al. 2014)を用いる。また、単語生成デコーダには、図6のように、一般的に用いられる入力データに対する注意機構(Wiseman et al. 2017; Chen et al. 2019)(図中では省略)だけでなく、内容選択モデルが予測した天気ラベルに対する注意機構を備える。これにより、テキスト生成において



重要な情報を明示的に記述するために導入する天気ラベルを考慮することが期待できる。

単語生成デコーダでは、タイムステップ  $t$  における単語  $w_t$  の生成確率は、下記の式により計算される：

$$p(w_t|w_{<t}, \mathbf{g}, \mathbf{a}, \mathbf{m}, \mathbf{l}) = \text{softmax}_{w_t}(W_s s_t^w), \quad (5)$$

$$s_t^w = \text{GRU}(w_{t-1}, s_{t-1}^w, c_t). \quad (6)$$

ここで、 $w_{t-1}$ 、 $s_{t-1}^w$  はタイムステップ  $t-1$  における出力単語および単語生成デコーダの内部状態を表す。 $W_s$  は重みパラメータである。また、タイムステップ  $t$  におけるベクトル  $c_t$  は、3 種類の入力データ  $(\mathbf{g}, \mathbf{a}, \mathbf{m})$  と天気ラベル  $\mathbf{l}$  それぞれの文脈ベクトル  $[c_t^g; c_t^a; c_t^m; c_t^l]$  を結合した文脈ベクトルである。例えばタイムステップ  $t$  における天気ラベル  $\mathbf{l}$  に対する文脈ベクトル  $c_t^l$  は、下記の式により導出される：

$$c_t^l = \sum_{i=1}^{|\mathcal{I}|} \alpha_{t,i}^l s_i^l, \quad (7)$$

$$\alpha_{t,i}^l = \frac{\exp(\eta(s_{t-1}^w, s_i^l))}{\sum_{j=1}^{|\mathcal{I}|} \exp(\eta(s_{t-1}^w, s_j^l))}. \quad (8)$$

ここで、 $s_i^l$  は天気ラベル  $l_i$  に対する内容選択モデルの内部状態ベクトルであり、 $\alpha_{t,i}^l$  は、 $t$  番目の出力単語と天気ラベル  $l_i$  に対する内容選択モデルの内部状態ベクトルのアライメントスコアを表す。式 8 におけるスコア関数  $\eta$  には、MLP を用いた。また、その他の文脈ベクトル  $c_t^g$ 、 $c_t^a$ 、 $c_t^m$  は、式 7、8 と同様に導出できることに注意されたい。

## 5 実験設定

### 5.1 データセット

実験には、株式会社ウェザーニューズ<sup>8</sup>のピンポイント天気サービスから収集された 2014 年から 2015 年の天気予報コメント 57,412 件を利用した。このうち、2014 年に配信された 28,555 件のコメントを学習データ、2015 年に配信された 14,464 件、14,393 件をそれぞれ開発データ、評価データとして使用した。また、数値予報マップのデータとして、京都大学生存圏研究所が運営する生存圏データベース<sup>9</sup>によって収集・配布されている気象庁作成の 2014 年から 2015 年の数値予報マップ 2,715 件を利用した。このうち、2014 年に配信された 1,344 件の数値予報マップを学習データ、1,326 件、1,329 件をそれぞれ開発データ、評価データとして利用している。ここで、これらの合計

<sup>8</sup> <https://weathernews.jp/>

<sup>9</sup> <http://database.rish.kyoto-u.ac.jp>

が数値予報マップの合計数である 2,715 件と一致していないが, これは, 開発および評価データの天気予報コメントは 2015 年のデータからサンプリングしており, 異なるエリアの数値予報マップを日本全体の 1 つのマップからそれぞれ抽出して使用しているためである. なお, 天気予報コメントとそれに対応する抽出された数値予報マップはエリアごとに固有のものであり, 開発用およびテスト用の天気予報コメントは重複していない. また, 数値予報マップと天気予報コメントの対応付けはそれぞれの配信時刻に基づき実施した. 具体的には, 天気予報コメントは既に配信された数値予報マップに基づき作成されるため, 天気予報コメントの配信日時より以前に配信された数値予報マップのうち, コメントの配信日時に直近の数値予報マップをコメントと対応付けしている.

天気予報コメントでは, 翌日までの天気について言及していることから, 24 時間先までの 3 時間ごとで合計 9 タイムステップからなる数値予報マップ  $g$  を使用した. また, AMeDAS により収集された観測データとして, 降水量, 気温, 風速, 日照時間の過去 24 時間の 10 分ごとの合計 144 タイムステップからなる観測値を使用した. 各エリアの観測データは, 天気予報コメントの対象エリアとそれぞれ 1 対 1 対応している. 天気予報コメントのメタ情報として, 配信日時 (月, 日, 曜日, 時刻) およびエリア名 (東京, 熊本 等) を使用した.

## 5.2 ハイパーパラメータ

本研究では数値予報マップ, 観測値データ, メタ情報の三種類の入力データを扱うため, それぞれに対するエンコーダが必要となる. 数値予報マップのエンコーダには, レイヤ数が 1 層の MLP と Bi-GRU を用い, 活性化関数は Tanh を使用した. また, 観測値データとメタ情報のエンコーダには, レイヤ数が 1 層の MLP を使用し, 活性化関数は Tanh とした. デコーダには, テキスト生成タスクにおいて一般的に用いられている RNNLM を使用した. RNN には, レイヤ数が 2 層の GRU を使用した. 単語埋め込みベクトルおよび数値予報マップのエンコーダの内部状態の次元数は 512, 観測値データおよびメタ情報のエンコーダの内部状態の次元数は 64, デコーダの内部状態の次元数は 512 とした.

ミニバッチのサイズは 200, 損失関数には交差エントロピー, モデルパラメータの最適化手法には Adam (Kingma and Ba 2015) を使用した. 学習率は 0.001, 学習時のエポック数は 25 に設定した. 推論時には, ビームサーチによりテキスト生成を行い, ビーム幅は 5 とした. 実験結果では, 学習における全 25 エポックのうち開発データに対するロスが 3 エポック連続で改善しない場合は早期終了し, その時点で開発データに対して最もロスが低いモデルを評価対象とし, 自動評価および人手評価の結果を報告する.

また, 近年の言語生成タスクでは Transformer (Vaswani et al. 2017) を用いた手法や事前学習モデル (Devlin et al. 2019; Kale and Rastogi 2020; Raffel et al. 2020) の導入が一般的となっている. しかし, 生成タスクの予備検証において Transformer ベースの手法との比較を実施した結果, RNN を上回る結果が得られなかったため, 提案モデルでは RNN ベースの手法を採用している. 加えて,

提案モデルでは、事前学習を行わない場合であっても一定の学習効果を得られることが分かったため、事前学習モデルの導入は行っていない。

### 5.3 評価指標

実験では、自動評価指標と人手評価により評価を実施した。実際に配信された天気予報コメントと生成テキストの単語の一致度合いを測る目的として、テキスト生成の研究で広く用いられている BLEU-4(Papineni et al. 2002) と ROUGE-1(Lin 2004) を使用した。これは、気象の専門家により記述され、実際に配信されている天気予報コメントには、晴れや雨といったユーザーにとって重要な情報が含まれていることが一般的であるためである。BLEU-4 の計測には SacreBLEU<sup>10</sup> (Post 2018)、ROUGE-1 の計測には rouge<sup>11</sup>を使用した。また、ROUGE-1 のスコアとして、 $F_1$  スコアを報告する。

しかしながら、生成テキストと参照テキストに含まれる情報が類似していたとしても語彙表現が異なる場合には、参照テキストとの単語の一致率に基づく BLEU や ROUGE といった自動評価指標では適切に評価できないことが懸念される。そこで、語彙表現が異なる場合でもテキストに含まれる情報の正確さを検証するために、参照テキストおよび生成テキストから表 2 に示す手がかり語に基づき抽出された天気ラベルの適合率 (*Precision*)、再現率 (*Recall*) およびそれらの調和平均である  $F_1$  スコアに基づく評価を実施した。これにより、生成テキストにおける内容選択の正確性を評価することができる。

人手評価では、一般ユーザーの視点から天気予報コメントを品質を評価するために、クラウドソーシングを用いて情報性、一貫性、文法性について、5 人の評価者により 3 段階の評価を実施した。表 3 に人手評価指標の概要を示す。情報性の評価では、システムの入力として使用した数値予報マップや気象観測データは使用せず、実際に配信された天気予報コメントを正解テキストとして提示し、それらと生成テキストを比較した上で評価を行うよう依頼した。これは、専門家ではない評価者にとって、これらの専門性の高いデータを理解した上で評価することは困難であると考えたためである。また、天気予報コメントの評価では、「晴れ」や「雨」といった天気予報のそれぞれの内容だけでなく、「晴れのち雨」などの天気の移り変わりについても正確に言及できているかを評価する必要がある。しかし、前述の天気ラベルに基づく内容選択の評価では、生成テキストから抽出された天気ラベルを独立に評価するため、天気の移り変わり（抽出された天気ラベルの順序）を考慮することが難しい。そのため、情報性の評価では、正解テキストと生成テキストを合わせて提示することで、天気予報の内容（晴れ、雨）だけでなく、天気の移り変わり（晴れのち雨）に関して評価に考慮されることを期待している。ここで、情報性に関する 3 段階の評価では、重要な情報の全てまたは一部だけ含まれている場合であっても内容に誤りがあれば 1 となることに注意され

<sup>10</sup> <https://github.com/mjpost/sacrebleu>

<sup>11</sup> <https://github.com/pltrdy/rouge>

評価指標	スコア	説明文
情報性	3	重要な情報を適切に言及できており, 理想的な天気予報コメントである.
	2	重要な情報が一部欠けているものの含まれている情報は適切であり, 概ね問題ない.
	1	含まれている情報が誤っており, 天気予報コメントとして不適切である.
一貫性	3	文章全体に一貫性があり, 読みづらさを全く感じない.
	2	文章の一貫性が欠けている箇所が一部あり, 読みづらさを感じる.
	1	文章全体の一貫性が欠けており, 内容を理解することが難しい.
文法性	3	文法的な誤りは全く含まれていない.
	2	文法的な誤りが一部含まれているが, 内容は理解することができる.
	1	文法的な誤りが多く含まれており, 内容を理解することが難しい.

表 3 人手評価指標

たい.

一貫性の評価では, 生成テキストにおける内容や文のつながりの自然さを評価対象としている. 例えば, 内容の一貫性が欠けている例として, 「今日は一日中晴れるため, 折りたたみ傘が必要です。」のように内容の辻褄が合わない場合が挙げられる. また, 文のつながりが不自然な例としては「午前は晴れます. そのため, 午後からは雨が降るのでご注意ください。」といった例が挙げられる.

人手評価には, 評価データからランダムに抽出した 40 件の天気予報コメントを使用した. なお, 各天気ラベルの有用性を検証するために, 各天気ラベルが紐づくコメントが 10 件以上になるように抽出を行っている. 各コメントに対して 5 人全ての評価者で評価を実施した. また, 人手評価で比較するモデル間の差の統計的有意性を検定するために, ウィルコクソンの符号順位検定 (Wilcoxon 1945) を用いた.

## 5.4 比較モデル

表 4 に実験に使用したモデルの概要を表す. 実験では, 数値予報マップのエンコード手法の検討として, CNN または MLP をエンコーダとしたモデル (2, 3) を比較する. また, 本研究で提案した天気ラベルを予測する内容選択モデルの有用性を確認するために, 内容選択モデルを導入したモデル (4) と導入しないモデル (2, 3) を比較する.

さらに, 生成テキストの情報性向上を目的に, Wang et al. (2020) が提案した内容一致制約損失 (*content-matching constraint loss*) を検証した. 内容一致制約とは, Data-to-Text タスクにおいて, 入力データと出力テキストに含まれる情報の類似性は高いという着想から, 入力データを表す状態ベクトルと出力テキストを表す埋め込みベクトルの距離を近づける制約である. Wang et al. (2020) は, 損失関数として二乗誤差を用いることにより, これら 2 つのベクトルの距離を近づける制約を提案し, 有用性を示した. 本研究では, Wang et al. (2020) の手法を参考に, 内容一致制約損失を

モデル	コンポーネント				
	エンコーダ	メタ情報	天気ラベル	内容一致制約	事例ベース推論
(1)	—	✓	—	—	✓
(2)	CNN	✓	—	—	—
(3)	MLP	✓	—	—	—
(4)	MLP	✓	予測	—	—
(5)	MLP	—	予測	✓	—
(6)	MLP	✓	予測	✓	—
(7)	MLP	✓	正解	✓	—

表 4 実験に使用したモデルの概要

提案モデルへ導入する．具体的には，言及すべき内容を選択する内容選択モデルの予測結果を生成テキストへ反映させることで生成テキストの情報性を向上させることを期待して，内容選択モデルの出力状態ベクトルと生成テキストを表す埋め込みベクトルの二乗誤差を計算する損失関数を導入した．ここで，内容一致制約を導入するモデルをモデル (6) とし，モデル (4) と比較することで有用性を検証する．また，実験では，メタ情報の導入による効果を確認するために，メタ情報埋め込みベクトルを用いないモデル (5) も合わせて比較した．

天気ラベルを導入するモデル (4, 5, 6) では，内容選択モデルの性能が生成テキストの品質に影響を与えることが考えられる．つまり，内容選択モデル自体の性能を改善することで，生成テキストのさらなる品質向上が期待できる．そこで，実験では，参照テキストから抽出した“正解”の天気ラベルを用いるモデル (7) を導入した．これにより，天気ラベルの導入により期待できる品質向上の上限を検証することができる．また，表 4 における「天気ラベル」列の「予測」，「正解」は，それぞれ分類器により予測された天気ラベル，または，参照テキストから抽出した正解の天気ラベルであることを表している．

また，その他の比較手法として，知識や経験に基づくルールベース手法 (Kukich 1983) や類似する過去の問題の解法に基づいて新たな問題の解法を類推する枠組みである事例ベース推論 (Adeyanju 2012) が考えられる．しかし，多数のルールやテンプレートの作成に専門知識と膨大な作業時間を要するルールベース手法を数値気象予報のシミュレーション結果のような複雑な数値データへ適用することは現実的ではない．一方，事例ベース推論は，現在の数値気象データに類似する過去の数値気象データの抽出ができれば，その過去の気象データに対する天気予報コメントを現在の気象データに対する天気予報コメントとして活用することが可能であると考えられる．そこで実験では，数値気象データへの適用がより実現性の高い事例ベース推論を比較手法として採用した．Adeyanju (2012) は，風に関する予報コメントの生成タスクにおいて，事例ベース推論を用いた手法を提案している．Adeyanju (2012) らの研究では，クエリとなる現在の風向きや風の強さを基に，事例デー

データベースから過去の類似事例を抽出し、現在のクエリに対する天気予報コメントを作成している。本研究では、Adeyanju (2012) の手法を参考に、数値予報マップ、気象観測値およびメタ情報をクエリとして事例データベースから類似事例を抽出し、現在のクエリに対する天気予報コメントを作成する。具体的には、まず、クエリである数値予報マップ、気象観測値、および、メタ情報<sup>12</sup>から1つの数値ベクトルを作成する。次に、事例データベースに含まれる事例についても同様に数値ベクトルを作成し、クエリとのコサイン類似度を計算する。最後に、現在のクエリと最も類似する過去の事例を抽出し、抽出された過去事例に紐づく天気予報コメントを現在のクエリに対する天気予報コメントとして採用する。事例データベースには、5.1 節の学習データを用いた。つまり、与えられた評価データに対する天気予報コメントを、上記の順に基づいて学習データの類似事例から抽出することになる。表4において、事例ベース推論に基づく手法をモデル (1) とする。

## 6 実験結果

表5に生成テキストと参照テキストの単語一致率に基づく BLEU および ROUGE スコア、および、生成テキストと参照テキストのそれぞれから抽出した天気ラベルに関する適合率 (P%)、再現率 (P%)、 $F_1$  スコア ( $F_1$ %) による評価結果を示す。表5では、各モデルの重みパラメータの初期値を変更させて、3 回実行した際の平均スコアを報告する。

### 6.1 数値予報マップのエンコード手法の比較

まず、数値予報マップのエンコード手法の検討として、CNN または MLP を数値予報マップのエンコーダとして採用したモデル (2)、(3) を比較する。表5の BLEU および ROUGE による自動評価において、MLP をエンコーダとするモデル (3) と CNN をエンコーダとして用いるモデル (2) で

モデル	BLEU	ROUGE	晴れ			雨			曇り			雪		
			P%	R%	$F_1$ %	P%	R%	$F_1$ %	P%	R%	$F_1$ %	P%	R%	$F_1$ %
(1)	7.7	40.0	76.2	73.4	74.8	75.5	67.5	71.3	55.6	48.6	51.9	47.3	77.6	58.7
(2)	12.7	42.8	83.5	67.6	74.7	72.8	83.6	77.8	58.5	59.8	59.0	75.2	50.1	60.2
(3)	13.0	43.5	83.2	68.4	74.9	74.6	83.5	78.8	59.8	60.3	59.9	75.7	53.3	62.3
(4)	12.9	43.8	81.0	78.5	79.7	78.6	80.0	79.3	62.5	55.9	58.9	75.9	60.4	67.2
(5)	12.7	43.2	81.2	78.9	79.9	78.0	84.0	80.9	60.8	66.1	63.1	80.0	55.2	65.1
(6)	13.2	43.9	81.0	78.4	79.7	76.6	84.1	80.2	60.6	59.3	59.8	77.7	58.5	66.6
(7)	14.6	45.5	94.9	84.5	89.4	84.4	92.9	88.4	84.7	85.6	85.1	91.3	63.8	75.1

表 5 各モデルの評価データに対する自動評価スコア

<sup>12</sup> One-hot エンコーディングによりメタ情報を数値ベクトル化した。

は、2つのエンコーダ間において顕著な差は見受けられなかった。一方、表5に示す、生成テキストから手がかり語により抽出した天気ラベルに対する適合率、再現率、 $F_1$ スコアの評価では、特に「雨」と「雪」において、モデル(3)がモデル(2)を上回るスコアが得られた。この結果から、MLPをエンコーダとして用いるモデル(3)は、CNNを用いるモデル(2)に比べて、天気予報コメントにおいて重要な情報である「雨」や「雪」について適切に言及できていることが推察できる。

以降の実験では、天気ラベルに基づく評価においてMLPを用いるモデル(3)がモデル(2)を上回っていることから、モデル(3)をベースモデルとし、各コンポーネントの有用性を検証する。

## 6.2 内容選択による生成テキストへの影響

次に、生成テキストの情報性向上を目的として導入した、内容選択モデルを検証するために、本コンポーネントを導入したモデル(4)と導入しないモデル(3)を比較する。表5に示す、生成テキストから手がかり語に基づき抽出した天気ラベルに関する $F_1$ スコアについて、内容選択モデルを導入したモデル(4)は、導入しないモデル(3)に比べてスコアが大きく向上していることを確認できた。特に、「晴れ」、「雪」の $F_1$ スコアについては約5%の改善が確認できており、これらの気象情報を生成テキストにおいて適切に言及できていることが推察できる。これらの結果から、本研究で導入した内容選択モデルのように、入力データから言及すべき内容を明示的に選択する外部機構を導入することで、生成テキストの情報性が改善することが確認できた。

また、表5の各天気ラベルに関する自動評価スコアは、生成テキストで各気象情報についてどの程度適切に言及できているかを表す指標であるが、提案モデルの単語生成デコーダは内容選択モデルの予測結果を参照するため、これらのスコアは内容選択モデルの精度に依存していると考えられる。つまり、内容選択モデルの精度が高い場合は生成テキストの情報性の向上を期待できるが、内容選択モデルの精度が低い場合は生成テキストへの悪影響が懸念される。そこで追加検証として、内容選択モデルの精度を評価し、その分類精度の生成テキストへの影響を調査した。表6に内容選択モデル単体の予測精度を示す。表6の内容選択モデルの精度と、表5に示すモデル(4)の生成テキストにおける内容選択の精度を比較すると、それぞれの予測精度は概ね匹敵していることが分かった。このことから、モデル(4)は内容選択モデルの予測結果を概ね反映できていることが推察

ラベル	P%	R%	$F_1$ %
晴れ	79.7	84.9	82.1
雨	79.9	80.5	80.2
曇り	61.5	62.5	61.6
雪	73.9	67.1	70.3

表6 評価データに対する内容選択モデルの予測精度

できる。しかし、現状では、モデル (4) の生成テキストにおける内容選択の精度は、内容選択モデル単体による予測精度よりも約 3% 程度低いことから、内容選択モデルの導入によりさらなる精度向上の余地があると考えられる。そのため、今後の課題の 1 つとして、内容選択モデルの予測結果を生成テキストへ十分に反映するための仕組みを導入することにより、生成テキストにおける内容選択の精度をさらに向上させることなどが考えられる。

### 6.3 内容一致制約の効果

内容選択モデルの予測結果と生成テキストを表す埋め込みベクトルの距離を近づけることで生成テキストの情報性が向上することを期待して導入した内容一致制約 (Wang et al. 2020) の効果について検証する。具体的には、内容一致制約を導入したモデル (6) と導入しないモデル (4) を比較する。

まず、表 5 の BLEU および ROUGE による自動評価では、内容一致制約を導入したモデル (6) は導入しないモデル (4) に対してわずかな改善に留まっていることが分かった。また、表 5 の内容選択に関する自動評価スコアでは、ほとんど改善が見受けられなかった。これらの結果から、内容一致制約による生成テキストの情報性改善の効果は限定的であることが推察できる。

### 6.4 メタ情報の導入による生成テキストへの影響

次に、生成モデルが天気予報コメントの配信される時間帯や対象エリアに依存する表現を用いてテキスト生成することを期待して導入したメタ情報埋め込みベクトルの効果について検証する。具体的には、メタ情報を用いたモデル (6) とそれを用いないモデル (5) の比較を行う。

表 5 の BLEU および ROUGE による自動評価では、メタ情報を導入したモデル (6) はそれを用いないモデル (5) と比べて、わずかにスコアが向上することを確認できた。しかしながら、生成テキストと参照テキストの単語の一致率に基づく BLEU や ROUGE といった自動評価スコアだけでは、モデル (6) がモデル (5) に比べてメタ情報に依存する表現を適切に生成できているか判断することが難しい。そこで、天気予報コメントの配信時刻やエリアなどのメタ情報に依存する表現について、参照テキストと比較して正しく出力できているか評価を実施した。具体的には、参照テキストに含まれるメタ情報の依存表現について、提案モデルより生成されたテキストにおいても適切に言及できているか評価する。

表 7 にメタ情報に依存する表現を参照テキストと比較して生成テキストにおいて適切に言及できているかを表す  $F_1$  スコア、適合率 (P%)、および、再現率 (R%) の自動評価結果を示す。表 7 における  $\Delta_{F_1}$  は、 $F_1$  スコアに関するモデル (6) とモデル (5) の差分を表している。また、表 8 に学習データ、開発データ、評価データの天気予報コメントにおける各依存表現の出現回数を示す。これらの評価対象として選定した依存表現は、開発データを参考に人手により抽出している。この結果によると、天気予報コメントのメタ情報として与えたモデル (6) はメタ情報を考慮しないモデル



カテゴリ	依存表現	モデル (5)			モデル (6)			$\Delta_{F_1}$
		P%	R%	$F_1\%$	P%	R%	$F_1\%$	
時刻に依存する表現	明日	93.7	88.7	91.1	97.8	92.7	95.1	+4.0
	今日	96.9	97.8	97.3	99.2	99.5	99.3	+2.0
	午前	10.5	1.3	2.3	9.0	1.5	2.6	+0.3
	午後	28.9	20.4	23.7	27.9	22.9	25.0	+1.3
曜日に依存する表現	( 月 )	0.0	0.0	0.0	28.5	31.2	29.3	+29.3
	( 火 )	0.0	0.0	0.0	25.2	35.2	29.2	+29.2
	( 水 )	2.9	0.7	1.1	22.1	40.9	28.4	+27.3
	( 木 )	19.1	3.3	5.6	21.5	26.1	23.5	+17.9
	( 金 )	0.6	0.4	0.4	13.5	27.2	18.0	+17.6
	( 土 )	0.0	0.0	0.0	8.0	19.2	11.3	+11.3
	( 日 )	0.0	0.0	0.0	21.8	37.4	27.4	+27.4
月に依存する表現	春	27.4	1.4	2.4	34.4	9.1	14.0	+11.6
	夏	25.5	8.2	12.4	26.5	15.1	19.1	+6.7
	秋	14.8	0.9	1.7	31.4	6.1	10.1	+8.4
	冬	17.6	3.9	6.2	17.7	4.7	7.3	+1.1
エリアに依存する表現	台風	47.1	17.8	23.8	31.8	13.1	16.3	-7.5
	高波	16.0	6.0	8.6	7.3	1.6	2.5	-6.1
	高潮	0.0	0.0	0.0	8.3	1.2	2.2	+2.2
	波	29.8	22.9	25.7	31.6	23.2	25.9	+0.2
	海	7.7	3.4	4.1	14.0	4.1	5.5	+1.4
	海岸	8.0	15.9	10.6	14.2	18.4	12.7	+2.1
	沿岸	5.6	3.3	4.2	2.9	1.2	1.7	-2.5
	マリン	8.0	2.4	3.7	45.3	3.4	6.3	+2.6
	山	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	山間	0.0	0.0	0.0	0.0	0.0	0.0	0.0

表 7 各依存表現に対する自動評価スコア

(5) に比べて、配信時刻に依存する「今日、明日」や、曜日に依存する「(月)、(火)」、月に依存する「春、夏」などの表現に対する  $F_1$  スコアが大きく向上していることを確認できた。同様に、エリアに依存する表現においても  $F_1$  スコアが全体的に向上していることが確認できた。しかし、評価データにおいて比較的出現回数が少ない依存表現（山、山間）については変化が見受けられなかった。また、選定した依存のうち、エリアに依存する表現である「高波、台風」について精度劣化が見受けられた。表 8 に示すように、この 2 つの依存表現は共起している回数が多い。そのため、メタ情報をモデルへ与えたことで台風または高波のいずれかを言及する精度が低下し、他方の依存表現に対する精度も劣化したことが考えられる。

以上の結果から、生成モデルへメタ情報を導入することにより、多くの事例において、メタ情報

カテゴリ	依存表現	各データにおける出現回数		
		学習データ	開発データ	評価データ
時刻に依存する表現	明日	9910	5176	5203
	今日	19328	9636	9521
	午前	1301	559	494
	午後	4628	2280	2314
曜日に依存する表現	( 月 )	229	82	91
	( 火 )	214	96	110
	( 水 )	238	130	101
	( 木 )	259	119	111
	( 金 )	305	93	92
	( 土 )	260	100	87
	( 日 )	220	75	73
月に依存する表現	春	795	279	319
	夏	835	596	590
	秋	529	428	442
	冬	1732	579	514
エリアに依存する表現	台風	488	277	255
	高波	186	159	129
	高潮	19	29	27
	波	275	214	220
	海	241	102	106
	海岸	94	65	67
	沿岸	219	99	81
	マリン	84	89	97
	山	106	27	35
	山間	22	3	3
(参考) 共起回数	(台風, 高波)	45	87	62

表 8 各依存表現の各データにおける出現回数

に依存した表現を適切に生成する性能が改善することが確認できた。一方で、いくつかの依存表現においては性能劣化が見受けられたことから今後の改善が求められる。

## 6.5 事例ベース推論に基づく手法との比較

エンコーダ・デコーダモデルに基づく提案モデルの有用性を検証するために、比較手法として導入した事例ベース推論に基づく手法 (Adeyanju 2012) と提案モデルの比較を行う。表 5 におけるモデル (1) が事例ベース推論に基づいたモデルの自動評価結果である。表 5 の BLEU および ROUGE に基づく自動評価において、提案モデルであるモデル (3) や (6) は、事例ベース推論に基づく手法であるモデル (1) よりも大幅に上回るスコアが得られた。この結果からエンコーダ・デコーダモデ

ルに基づく提案モデルは、事例ベース推論に基づくモデル (1) よりも参照テキストに類似した天気予報コメントを生成できていることが推察できる。一方で、生成テキストにおける内容選択の正確性を表す各天気ラベルの精度においては、事例ベース推論に基づくモデル (1) は提案モデルであるモデル (3) や (6) よりもスコアが劣ってはいるものの、比較的高いスコアを得られていることが分かった。つまり、事例ベース推論に基づくモデル (1) の生成テキストは、提案モデルと比べて参照テキストとの類似性はやや劣るものの、適切な内容選択ができていることが推察できる。

上記の結果から、数値予報マップ、気象観測値、メタ情報をクエリとした事例ベース推論に基づくモデル (1) では、最も類似度が高い事例を事例データベースから抽出することで、内容選択の正確性が高い天気予報コメントが得られることが分かった。このことを踏まえると、最も類似度が低い事例が抽出された場合には、生成テキストにおける内容選択の正確性は低くなることが予想される。そこで追加検証として、クエリに類似した事例を事例データベースから抽出する際に用いる類似度と生成テキストにおける内容選択の正確性の関係性を調査した。具体的には、クエリに類似した事例を事例データベースから抽出する際に、最も類似度が高い事例を抽出する場合のモデル  $(1)_{max}$  と最も類似度が低い事例を抽出する場合のモデル  $(1)_{min}$  を比較する。ここで、最も類似度が高い事例を抽出した場合のモデル  $(1)_{max}$  は、表 5 におけるモデル (1) と同じモデルであることに注意されたい。表 9 に各モデルの自動評価結果を示す。これら 2 つのモデルを比較すると、クエリに最も類似度が高い事例を事例データベースから抽出した場合のモデル  $(1)_{max}$  は、クエリに最も類似度が低い事例を抽出した場合のモデル  $(1)_{min}$  よりも大幅に自動評価スコアが向上していることが分かった。これらの結果からも、事例データベース推論に基づく手法において、最も類似度が高い事例を事例データベースから抽出することで、内容選択の正確性が高い天気予報コメントを得られることが確認できた。

また、表 10 に事例ベース推論に基づくモデル  $(1)_{max}$  およびモデル  $(1)_{min}$  による生成テキストの例を示す。表 10 における「正解」は、評価データにおける一事例であり、2015 年 7 月 23 日 (木曜) 午前 0 時 22 分に実際に配信された天気予報コメントである。モデル  $(1)_{max}$  およびモデル  $(1)_{min}$  のそれぞれの生成テキストは、数値予報マップ、気象観測値、メタ情報からなるクエリの類似度に基づいて事例データベースから抽出された天気予報コメントである。まず、クエリに最も類似度が高い事例を用いたモデル  $(1)_{max}$  の生成テキストは、正解テキストと同様に、雨と曇り空について言

モデル	BLEU	ROUGE	晴れ			雨			曇り			雪		
			P%	R%	F <sub>1</sub> %	P%	R%	F <sub>1</sub> %	P%	R%	F <sub>1</sub> %	P%	R%	F <sub>1</sub> %
$(1)_{max}$	7.7	40.0	76.2	73.4	74.8	75.5	67.5	71.3	55.6	48.6	51.9	47.3	77.6	58.7
$(1)_{min}$	4.4	34.6	49.8	53.5	51.6	48.5	36.4	41.6	47.0	40.0	43.2	0.0	0.0	0.0

表 9 事例ベース推論に基づくモデルの評価データに対する自動評価スコア

及できていることが確認できた. 一方, クエリに最も類似度が低い事例を用いたモデル  $(1)_{min}$  の生成テキストでは, 「今日は冬晴れの一日。お出かけ日和」のように, 正解テキストとは異なった内容が言及されている. これは, モデル  $(1)_{min}$  では, 正解の気象データと最も類似度が低い事例の天気予報コメントを生成テキストとして用いるため, 言及される内容は必然的に正解テキストと大きく異なることが要因となっている. 以上の結果からも, 事例ベース推論に基づく手法において, 事例データベースから抽出する際の類似度によって生成テキストで言及される内容およびその正確性が変化することが確認できた.

## 6.6 人手評価結果

表 11 に人手評価の結果を示す. 表 11 における,  $\dagger$  は統計的に有意差があることを示す ( $p < 0.05$ ). また, 「事例数」列は, 人手評価データ 40 件のうち, 各天気ラベルを含む事例の件数を表している. ここで, 各天気予報コメントには図 1 の事例のように 1 つ以上の天気ラベルが含まれることに注意されたい.

人手評価結果によると, 本研究で提案した天気ラベルの予測タスクにより明示的な内容選択を行うモデル (6) は, 明示的な内容選択を行わないモデル (3) に比べて, 統計的に有意に情報性が優れていることが分かった. これは, 表 5 の自動評価結果から明らかになったように, 提案モデル (6) は

モデル	生成テキスト (配信日時: 2015 年 7 月 23 日 (木曜) 午前 0 時 22 分)
正解	今日は雨が降りやすく、お出かけには傘が必須となりそうです。日差しはほとんどなくても湿気たっぷりの空気でムシムシと感ずる一日に。
$(1)_{max}$	今日の朝は雨が降りやすく傘を持ってお出かけをして下さい。昼間は雲が多くスッキリとしない空模様。帰宅の時には傘の置き忘れに注意して下さい。
$(1)_{min}$	今日は冬晴れの一日。お出かけ日和ですが、冷たい風がピューッと吹くので、暖かくしてお出かけください。

表 10 事例ベース推論に基づく手法の生成テキスト

天気ラベル	モデル (3)			モデル (6)			事例数
	情報性	一貫性	文法性	情報性	一貫性	文法性	
晴れ	1.92	2.91	2.91	2.10	2.82	2.88	26
雨	2.02	2.93	2.92	2.13	2.88	2.90	26
曇り	1.99	2.93	2.94	2.12	2.83	2.89	19
雪	1.88	2.95	2.92	1.95	2.91	2.94	13
全ラベル	1.98	2.92 $^\dagger$	2.92	2.10 $^\dagger$	2.86	2.90	40

表 11 人手評価の結果

モデル (3) よりも生成テキストにおける内容選択の正確性が高いことが要因として考えられる。

一方で、文章全体の一貫性の観点においては、両モデルのスコアは十分に高いものの、モデル (6) はモデル (3) よりも劣ることが分かった。これは、生成テキストに含まれる情報量が増えたことで、それらの一貫性を担保することが難しくなったことが要因として考えられる。この問題の対策としては、テキストで言及する内容を予測する内容選択だけではなく、それらをどのような順序で言及するかを考慮するための内容プランニングを明示的に実施することが挙げられる (Wiseman et al. 2017; Iso et al. 2019)。また、モデル (6) の生成テキストでは、以下に示す事例 (e) のように文と文のつながりの不自然さに起因して一貫性スコアが低く評価された事例<sup>13</sup>も確認された。この例では、1 文目と 2 文目を繋ぐ接続詞として「ただ」が使われていることで、1 文目から推察できる内容（段々と日差しが届いて晴れるため、午後から過ごしやすくなる事）に対して、さらに 2 文目で補足している状況となり、文のつながりに不自然さが生じていたことが原因として考えられる。

- (e) 今日は朝まで雨が降りますが、段々と日差しが届いてお出かけ日和になります。ただ、昼間は過ごしやすい体感になりそうです。

表 12, 13, 14 に各モデルの出力例と実際に配信された正解の天気予報コメントの例を示す。表

モデル	I	C	G	生成テキスト（配信日時: 2015 年 6 月 22 日（月曜）午前 0 時 2 分）
正解	—	—	—	今日は、うっすら雲が広がりやすいものの、日差しが届きます。夏至の日差しは強烈なので、紫外線・暑さ対策が欠かせません。
(3)	2.0	3.0	3.0	今日（月）は日差しが届きますが、段々と雲が広がります。 <u>午後はニワカ雨の可能性があるので、折りたたみ傘があると安心です。</u>
(6)	2.8	2.8	3.0	今日（月）は雲が広がりやすいものの、日差しが届く時間もあります。ムシムシとした暑さになるので、熱中症対策を忘れずに。

表 12 「豊橋」エリアに配信された天気予報コメントと各モデルの生成テキスト

モデル	I	C	G	生成テキスト（配信日時: 2015 年 12 月 26 日（土曜）午前 0 時 0 分）
正解	—	—	—	今日は日差しが届いても、ゆっくり天気は下り坂になります。夕方から雪が降り出すので、外出は早めがオススメ。昨日より冷え込みが増すので、防寒が必須です。
(3)	2.0	3.0	2.6	今日（土）は日差しが届いても、ニワカ雨の可能性があり。お出かけには折りたたみ傘があると安心です。昼間でも気温は低空飛行で体を温めるのも良さそうです。
(6)	2.4	3.0	3.0	今日（土）は日差しが届いても雲が優勢の天気。ニワカ雪の可能性があるので、折りたたみ傘があると安心です。また、寒さが続くので防寒対策を万全にしてください。

表 13 「白糠」エリアに配信された天気予報コメントと各モデルの生成テキスト

<sup>13</sup> 一貫性の平均スコアは 1.8.

モデル	I	C	G	生成テキスト（配信日時: 2015 年 7 月 18 日（土曜）午前 10 時 46 分）
正解	—	—	—	今日も変わりやすい天気です。急にザッと雨が降る可能性があるので、日差しがあっても油断できません。ムシッとする風もやや強いままです。
(3)	1.8	2.8	2.8	今日は雲が広がりやすいものの、日差しが届く時間もあります。この時期の紫外線は強烈なので、UV対策を忘れずにお出かけ下さい。
(6)	2.6	2.8	2.8	今日は変わりやすい天気。日差しが届いても、急な雨もあります。モクモクした雲が近づいて来たら天気急変のサインです。

表 14 「東京」エリアに配信された天気予報コメントと各モデルの生成テキスト

12, 13, 14 における I, C, G はそれぞれ、情報性 (*Informativeness*), 一貫性 (*Consistency*), 文法性 (*Grammaticality*) に関する評価者 5 人の平均スコアを表す。表 12 では、各モデルの生成テキストにおいて正解テキストと同様に雲について言及できているものの、モデル (3) の生成テキストでは正解テキストでは言及されていない雨について記述されている。この結果から、モデル (3) の生成テキストは誤った情報が含まれていると判断され、モデル (6) よりも情報性のスコアが低く付けられたと考えられる。表 13 では、各モデルの生成テキストでは日差しや気温の低下について言及できている。一方、モデル (6) は正解テキストと同様に雪についても言及できているが、モデル (3) では雪ではなく、ニワカ雨と言及している。このことから、表 13 におけるモデル (3) の生成テキストは、情報性が低いと判断されたことが考えられる。表 14 では、モデル (6) は正解テキストと同様に日差しや雨について言及できているため比較的高い情報性のスコアが付けられている。一方、モデル (3) では、日差しについては言及できているものの、雨についての記述は含まれていない。そのため、モデル (3) の生成テキストはモデル (6) よりも情報性が低いと判断されたことが考えられる。

また、情報性の評価では、正解テキストと生成テキストを比較して評価を行っているが、正解テキストの何を重要な情報とみなすかは評価者ごとに結果が分かれる可能性がある。例えば、以下の正解テキスト (f) と生成テキスト (g) では、生成テキストに対して評価者 5 名のうち 3 名からは 3, 2 名からは 2 の評価結果が得られた。この事例において、正解テキストと比較して生成テキストで言及されていない差分となる情報は、風や気温に関する情報であるため、2 名の評価者は風や気温に関する情報も重要な情報とみなし、生成テキストに対して 2 を付与したことが推察できる。

- (f) 今日（木）は晴れたり曇ったり。ニワカ雨の可能性もあるので、お出かけの際は折りたたみ傘があると安心です。風が吹くと涼しい～肌寒いくらいの体感になります。
- (g) 今日（木）は日差しが届いても、ニワカ雨の可能性がります。お出かけの際は折りたたみ傘があると安心です。

## 7 おわりに

本研究では、数値気象予報のシミュレーション結果から天気予報コメントを生成するための Data-to-Text モデルを提案した。天気予報コメント生成タスクには、(i) 複数の物理量からなる時系列数値データの考慮する必要がある、(ii) コメントが書かれる時間帯や対象エリアに依存した表現が用いられる、(iii) 天気予報コメントにおいて情報性が重要である、といった課題があり、本研究ではそれぞれの課題に対して手法を提案した。実験では、自動評価および人手評価によって提案手法の有用性を示した。

今後の課題として2点挙げる。まず、1点目の課題として、人手評価において、提案モデルの生成テキストの情報性の向上は見受けられたものの、わずかな劣化が観測された一貫性の課題が挙げられる。この課題の解決策として、生成テキストの情報性向上を目的とした、入力データからの内容選択だけではなく、それらの内容をどのような順序で言及するかを考慮するための内容プランニングを導入する方法などが考えられる。

2点目の課題として、生成モデルの入力データとして使用した気象データの入力形式の検討が挙げられる。本研究では、入力データとして、数値気象予報のシミュレーション結果の生データである数値予報マップを使用した。これらのデータを基に作成された SUMTIME-METEO や WEATHERGOV などの構造化データを用いる方法も考えられる。これらの構造化データは、数値予報マップといった生データに比べて情報量が落ちることが懸念されるものの、人手や機械により情報が整理された上で構造化されているため、重要な情報を選定する内容選択の問題がより簡単になる可能性が考えられる。また、その他にも、生データや構造化データのいずれかだけでなく、両者を組み合わせた入力形式も考えられる。そのため、今後の課題として、天気予報コメントの生成タスクにより適した気象データの入力形式やそれらの組み合わせを検討したい。

## 謝 辞

本論文の一部は、The 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL2021) で発表したものです (Murakami et al. 2021)。

## 参考文献

- Adeyanju, I. (2012). “Generating Weather Forecast Texts with Case Based Reasoning.” *International Journal of Computer Applications*, **45**, pp. 35–40.
- Angeli, G., Liang, P., and Klein, D. (2010). “A Simple Domain-Independent Probabilistic Approach to Generation.” In *Proceedings of the 2010 Conference on Empirical Methods in*

- Natural Language Processing*, pp. 502–512. Association for Computational Linguistics.
- Aoki, K., Miyazawa, A., Ishigaki, T., Aoki, T., Noji, H., Goshima, K., Kobayashi, I., Takamura, H., and Miyao, Y. (2019). “Controlling Contents in Data-to-Document Generation with Human-Designed Topic Labels.” In *Proceedings of the 12th International Conference on Natural Language Generation*, pp. 323–332. Association for Computational Linguistics.
- Aoki, T., Miyazawa, A., Ishigaki, T., Goshima, K., Aoki, K., Kobayashi, I., Takamura, H., and Miyao, Y. (2018). “Generating Market Comments Referring to External Resources.” In *Proceedings of the 11th International Conference on Natural Language Generation*, pp. 135–139. Association for Computational Linguistics.
- Bahdanau, D., Cho, K., and Bengio, Y. (2015). “Neural Machine Translation by Jointly Learning to Align and Translate.” In *Proceedings of the 3rd International Conference on Learning Representations*.
- Barzilay, R. and Lapata, M. (2005). “Collective Content Selection for Concept-to-Text Generation.” In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pp. 331–338. Association for Computational Linguistics.
- Belz, A. (2007). “Probabilistic Generation of Weather Forecast Texts.” In *Proceedings of the 2007 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 164–171. Association for Computational Linguistics.
- Chen, S., Wang, J., Feng, X., Jiang, F., Qin, B., and Lin, C.-Y. (2019). “Enhancing Neural Data-To-Text Generation Models with External Background Knowledge.” In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pp. 3022–3032. Association for Computational Linguistics.
- Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014). “Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation.” In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1724–1734. Association for Computational Linguistics.
- Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. (2014). “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling.” In *NIPS 2014 Workshop on Deep Learning, December 2014*.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*:



- Human Language Technologies, Volume 1*, pp. 4171–4186. Association for Computational Linguistics.
- Duboue, P. A. and McKeown, K. R. (2001). “Empirically Estimating Order Constraints for Content Planning in Generation.” In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, pp. 172–179, Toulouse, France. Association for Computational Linguistics.
- Duboue, P. A. and McKeown, K. R. (2003). “Statistical Acquisition of Content Selection Rules for Natural Language Generation.” In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, pp. 121–128.
- Gatt, A. and Krahmer, E. (2018). “Survey of the State of the Art in Natural Language Generation: Core Tasks, Applications and Evaluation.” *Journal of Artificial Intelligence Research*, **61** (1), pp. 65–170.
- Goldberg, E., Driedger, N., and Kittredge, R. I. (1994). “Using Natural-Language Processing to Produce Weather Forecasts.” *IEEE Expert*, **9** (2), pp. 45–53.
- Iso, H., Uehara, Y., Ishigaki, T., Noji, H., Aramaki, E., Kobayashi, I., Miyao, Y., Okazaki, N., and Takamura, H. (2019). “Learning to Select, Track, and Generate for Data-to-Text.” In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 2102–2113. Association for Computational Linguistics.
- Kale, M. and Rastogi, A. (2020). “Text-to-Text Pre-Training for Data-to-Text Tasks.” In *Proceedings of the 13th International Conference on Natural Language Generation*, pp. 97–102, Dublin, Ireland. Association for Computational Linguistics.
- Kerpedjiev, S. M. (1992). “Automatic Generation of Multimodal Weather Reports from Datasets.” In *3rd Conference on Applied Natural Language Processing*, pp. 48–55. Association for Computational Linguistics.
- Kingma, D. P. and Ba, J. (2015). “Adam: A Method for Stochastic Optimization.” In *Proceedings of the 3rd International Conference on Learning Representations*.
- Kittredge, R., Polguere, A., and Goldberg, E. (1986). “Synthesizing Weather Forecasts From Formatted Data.” In *Proceedings of the 11th International Conference on Computational Linguistics*.
- Kukich, K. (1983). “Design of A Knowledge-Based Report Generator.” In *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, pp. 145–150. Association for Computational Linguistics.
- Lebet, R., Grangier, D., and Auli, M. (2016). “Neural Text Generation from Structured Data with Application to the Biography Domain.” In *Proceedings of the 2016 Conference on*

- Empirical Methods in Natural Language Processing*, pp. 1203–1213. Association for Computational Linguistics.
- Li, J., Galley, M., Brockett, C., Spithourakis, G., Gao, J., and Dolan, B. (2016). “A Persona-Based Neural Conversation Model.” In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 994–1003. Association for Computational Linguistics.
- Liang, P., Jordan, M., and Klein, D. (2009). “Learning Semantic Correspondences with Less Supervision.” In *Proceedings of Association for Computational Linguistics and International Joint Conference on Natural Language Processing*, pp. 91–99. Association for Computational Linguistics.
- Lin, C.-Y. (2004). “ROUGE: A Package for Automatic Evaluation of Summaries.” In *Text Summarization Branches Out*, pp. 74–81, Barcelona, Spain. Association for Computational Linguistics.
- Liu, T., Wang, K., Sha, L., Chang, B., and Sui, Z. (2018). “Table-to-Text Generation by Structure-Aware Seq2seq Learning.” In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, pp. 4881–4888.
- Long, X., Gan, C., and de Melo, G. (2018). “Video Captioning with Multi-Faceted Attention.” *Transactions of the Association for Computational Linguistics*, **6**, pp. 173–184.
- Luong, T., Sutskever, I., Le, Q., Vinyals, O., and Zaremba, W. (2015). “Addressing the Rare Word Problem in Neural Machine Translation.” In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 11–19. Association for Computational Linguistics.
- Ma, S., Yang, P., Liu, T., Li, P., Zhou, J., and Sun, X. (2019). “Key Fact as Pivot: A Two-Stage Model for Low Resource Table-to-Text Generation.” In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 2047–2057. Association for Computational Linguistics.
- Mei, H., Bansal, M., and Walter, M. R. (2016). “What to Talk about and How? Selective Generation using LSTMs with Coarse-to-Fine Alignment.” In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 720–730. Association for Computational Linguistics.
- Mikolov, T., Karafiát, M., Burget, L., Černocký, J., and Khudanpur, S. (2010). “Recurrent Neural Network Based Language Model.” In *Proceedings of the 11th Annual Conference of the International Speech Communication Association*, pp. 1045–1048. International Speech

Communication Association.

- Murakami, S., Tanaka, S., Hangyo, M., Kamigaito, H., Funakoshi, K., Takamura, H., and Okumura, M. (2021). “Generating Weather Comments from Meteorological Simulations.” In *Proceedings of The 16th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 1462–1473. Association for Computational Linguistics.
- Murakami, S., Watanabe, A., Miyazawa, A., Goshima, K., Yanase, T., Takamura, H., and Miyao, Y. (2017). “Learning to Generate Market Comments from Stock Prices.” In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1374–1384. Association for Computational Linguistics.
- Novikova, J., Dušek, O., and Rieser, V. (2017). “The E2E Dataset: New Challenges for End-to-End Generation.” In *Proceedings of the 18th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, Saarbrücken, Germany. arXiv:1706.09254.
- Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). “BLEU: A Method for Automatic Evaluation of Machine Translation.” In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 311–318. Association for Computational Linguistics.
- Post, M. (2018). “A Call for Clarity in Reporting BLEU Scores.” In *Proceedings of the 3rd Conference on Machine Translation: Research Papers*, pp. 186–191, Belgium, Brussels. Association for Computational Linguistics.
- Puduppully, R., Dong, L., and Lapata, M. (2019). “Data-to-Text Generation with Content Selection and Planning.” In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, pp. 6908–6915.
- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J. (2020). “Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer.” *Journal of Machine Learning Research*, **21** (140), pp. 1–67.
- Reiter, E. and Dale, R. (1997). “Building Applied Natural Language Generation Systems.” *Natural Language Engineering*, **3** (1), pp. 57–87.
- Reiter, E., Sripada, S., Hunter, J., Yu, J., and Davy, I. (2005). “Choosing Words in Computer-Generated Weather Forecasts.” *Artificial Intelligence*, **167** (1-2), pp. 137–169.
- Sha, L., Mou, L., Liu, T., Poupart, P., Li, S., Chang, B., and Sui, Z. (2018). “Order-Planning Neural Text Generation From Structured Data.” In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, pp. 5414–5421.
- Sripada, S., Reiter, E., and Davy, I. (2004). “SumTime-Mousam: Configurable Marine Weather Forecast Generator.” *Expert Update*, **6**, pp. 4–10.

- Sripada, S., Reiter, E., Hunter, J., and Yu, J. (2003). “Exploiting a Parallel TEXT - DATA corpus.” In *Proceedings of Corpus Linguistics*, pp. 734–743.
- Sripada, S. G., Reiter, E., Hunter, J., Yu, J., and Davy, I. P. (2002). “Modelling the Task of Summarising Time Series Data Using KA Techniques.” In Macintosh, A., Moulton, M., and Preece, A. (Eds.), *Applications and Innovations in Intelligent Systems IX*, pp. 183–196, London. Springer London.
- Sutskever, I., Vinyals, O., and Le, Q. V. (2014). “Sequence to Sequence Learning with Neural Networks.” In *Proceedings of the 27th International Conference on Neural Information Processing Systems*, pp. 3104–3112.
- Tu, Z., Lu, Z., Liu, Y., Liu, X., and Li, H. (2016). “Modeling Coverage for Neural Machine Translation.” In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 76–85, Berlin, Germany. Association for Computational Linguistics.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). “Attention Is All You Need.” In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (Eds.), *Advances in Neural Information Processing Systems 30*, pp. 5998–6008. Curran Associates, Inc.
- Wang, Z., Wang, X., An, B., Yu, D., and Chen, C. (2020). “Towards Faithful Neural Table-to-Text Generation with Content-Matching Constraints.” In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 1072–1086. Association for Computational Linguistics.
- Wilcoxon, F. (1945). “Individual Comparisons by Ranking Methods.” *Biometrics Bulletin*, **1** (6), pp. 80–83.
- Wiseman, S., Shieber, S., and Rush, A. (2017). “Challenges in Data-to-Document Generation.” In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 2253–2263. Association for Computational Linguistics.
- Yao, L., Torabi, A., Cho, K., Ballas, N., Pal, C., Larochelle, H., and Courville, A. (2015). “Describing Videos by Exploiting Temporal Structure.” In *Proceedings of the 2015 IEEE International Conference on Computer Vision*, pp. 4507–4515. IEEE Computer Society.

## 略歴

村上聡一郎：2015年熊本高等専門学校専攻科電子情報システム工学専攻修了。2017年東京工業大学大学院博士前期課程修了。2019年より、東京工業大学工学院博士後期課程に在籍。自然言語処理、特に自然言語生成に関する研究開発に従事。

言語処理学会会員.

**田中 天**: 2016 年大阪大学工学部電子情報学科修了. 2019 年東京工業大学大学院博士前期課程修了.

**萩行 正嗣**: 2014 年京都大学大学院情報学研究科知能情報学専攻博士後期課程修了. 博士 (情報学). 同年より現職. 自然言語処理の研究に従事. 言語処理学会会員, 情報処理学会会員. AI 防災協議会常務理事.

**上垣外英剛**: 2012 年東京工業大学情報工学科卒業. 2014 年東京工業大学大学院総合理工学研究科知能システム科学専攻博士前期課程修了. 2017 年東京工業大学大学院 総合理工学研究科知能システム科学専攻博士後期課程修了. 博士 (工学). 同年 NTT コミュニケーション科学基礎研究所入所. 2018 年より東京工業大学科学技術創成研究院未来産業技術研究所助教. 自然言語処理分野の研究に従事. 情報処理学会, 言語処理学会, ACL 各会員.

**船越孝太郎**: 2005 年東京工業大学大学院情報理工学研究科計算工学専攻博士後期課程修了. 博士 (工学). 2006 年より (株) ホンダ・リサーチ・インスティテュート・ジャパン リサーチャー. 2017 年より京都大学大学院情報学研究科知能情報学専攻特定准教授. 2020 年より東京工業大学科学技術創成研究院未来産業技術研究所准教授. 自然言語処理およびマルチモーダル対話システムに関する研究に従事. 言語処理学会, 人工知能学会, 情報処理学会各会員.

**高村 大也**: 1997 年東京大学工学部計数工学科卒業. 2000 年同大大学院工学系研究科計数工学専攻修了 (1999 年はオーストリアウィーン工科大学にて研究). 2003 年奈良先端科学技術大学院大学情報科学研究科博士課程修了. 博士 (工学). 2003 年から 2010 年まで東京工業大学精密工学研究所助教. 2006 年にはイリノイ大学にて客員研究員. 2010 年から 2016 年まで同准教授. 2017 年から 2021 年 3 月まで同教授. 2017 年から産業技術総合研究所人工知能センター知識情報研究チーム研究チーム長.

**奥村 学**: 1962 年生. 1984 年東京工業大学工学部情報工学科卒業. 1989 年同大学院博士課程修了. 同年, 東京工業大学工学部情報工学科助手. 1992 年北陸先端科学技術大学院大学情報科学研究科助教授, 2000 年東京工業大学精密工学研究所助教授, 2009 年同教授. 現在は, 科学技術創成研究院教授. 2017 年より, 理化学研究所革新知能統合研究センター (AIP) 客員研究員を兼務. 工学博士.

(2021 年 4 月 30 日 受付)

(2021 年 7 月 31 日 再受付)

(2021 年 9 月 6 日 採録)