論文 / 著書情報
Article / Book Information

題目(和文)	Data-to-Textモデルの高度化に関する研究
Title(English)	
著者(和文)	村上聡一朗
Author(English)	Soichiro Murakami
出典(和文)	学位:博士（工学）, 学位授与機関:東京工業大学, 報告番号:甲第12190号, 授与年月日:2022年9月22日, 学位の種別:課程博士, 審査員:奥村 学,熊澤 逸夫,中山 実,篠﨑 隆宏,船越 孝太郎,白井 清昭
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第12190号, Conferred date:2022/9/22, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Data-to-Text モデルの高度化に関する研究

東京工業大学
工学院 情報通信系
情報通信コース

博士論文

指導教員 教授 奥村 学
村上 聡一朗

2022 年 8 月

目次

図目次

表目次

第 1 章

序論

1.1 研究背景

インターネットやセンサ等の情報基盤技術の普及に伴い，金融や医療，交通などの幅広い分野において，様々なデータが日々大量に蓄積されている．これらのデータを有効活用する手段の一つとして，グラフ等を用いた可視化などが考えられる．しかし，大規模かつ多種多様なデータを専門知識のない人が見て解釈することは容易ではなく，専門家であったとしてもデータから重要な情報を読み取るためには時間がかかる．そのため，データの概要や重要な情報を人間にとって分かりやすく説明するための自然言語生成技術に注目が集まっている (Gatt et al., 2018).

Data-to-Text 生成課題は，表データや数値データなど様々な形式のデータからそれを説明するテキストを生成する研究課題であり，これまで多くのドメインで研究が行われてきた．例えば，バスケットボールの試合要約 (Iso et al., 2019; Puduppully et al., 2019; Wiseman et al., 2017) やレストラン紹介文 (Novikova et al., 2017)，株価の市況コメント (K. Aoki and Kobayashi, 2016; Kukich, 1983)，天気予報コメント (Angeli et al., 2010; Belz, 2007) など，その対象のドメインは多岐にわたる．また，Data-to-Text 生成技術として，伝統的には数多くのルールとテンプレートを組み合わせたルールベース手法が中心的であったが，近年では入力データと出力テキストの対応関係からテキスト生成規則を自動的に獲得する機械学習ベースの手法が主流となっている (Gatt et al., 2018)．特に最近では，機械翻訳や文書要約でも広く用いられているニューラル言語生成技術に基づく手法が標準的手法となっている (Sharma et al., 2022).

ニューラル言語生成技術に基づく手法は，人間が書いたテキストと遜色ないほど流暢なテキストが生成できるとして注目を集めている．しかし，その一方でこの手法は生成時の制御が難しく，入力データの事実と異なる内容や重要性の低い内容を含むテキストが生成される課題も数多く指摘されている (Tian et al., 2019; Z. Wang et al., 2020; Wiseman et al., 2017)．これは特に生成テキストの内容の正確さが求められる様々なドメインにおいて Data-to-Text 生成技術を実用化するという観点において，重要な課題といえる．例えば，株価から市況コメントを生成するシステムにおいて，人間が書いたような流暢な市況コメントが生成されていたとしても，言及される株価やその値動き等の内容が事実と異なれば，その生成システムや市況コメントは実用的とはいえない.

この課題に対して，入力データに忠実なテキストの生成に向けた様々な研究が取り組まれている (W. Li et al., 2022)．その中の研究の方向性の一つとして，モデル自体を高度化する取り組みがある．この方向性では，入力データの特徴を適切に捉えるようモデルを設計し，正確にテキスト化することを目指す．例えば，表データからテキストを生成する Table-to-Text タスクの研究では，表データ中の固有名詞や数値を直接参照するためのコピー機構 (Wiseman et al., 2017) や言及すべき重要な情報を予測する内容選択モデル (Ma et al., 2019) 等の生成モデルを補助する外部機構を導入することで，生成テキストの品質向上に寄与することが報告されている．一方で，これらの外部機構は表データとテキストのように，入力データと出力テキストが単語の表層に基づいて部分的に対応付くことを前提とした手法であり，表データ以外の入力を対象としたタスクにそのまま適用することが難しいという側面もある．実世界の様々なデータを対象とする Data-to-Text 生成課題では，表データだけでなく数値や画像のように様々な形式のデータを扱うことから，対象タスクのデータに対して適切にモデルや外部機構を設計することが求められている．

1.2 本研究の貢献

本研究では，Data-to-Text 生成課題における生成テキストの正確性の問題に対して，モデル自体の高度化を通して貢献する．そのための方向性として，これまでの研究で有用性が示されてきたコピー機構や内容選択モデルといった外部機構を生成モデルに導入する方法に着目する．

本研究では，Data-to-Text 生成技術の実用化を見据え，実世界の様々な場面で観測される時系列数値データを対象とするタスクに焦点をあてる．その中でもデータの特徴が異なるドメインとして，株価の市況コメントおよび天気予報コメントに着目する．各タスクを対象とした Data-to-Text モデルの入力として，それぞれのテキストが記述される際に実際に参照される数値データである，時系列株価データおよび数値気象予報を用いる．本研究では各タスクに対して，Data-to-Text モデルの高度化を通して数値データの特徴を捉え正確にテキスト化するための手法を探求する．

以降では各タスクにおける課題とそれに対する本研究の貢献を整理する．

1.2.1 時系列株価データからの市況コメントの自動生成

本研究では，日経平均株価の時系列株価データから市況コメントを自動生成するタスクに取り組む．本研究では，株価の市況コメントにおける 3 つの特徴に着目する．1 つ目は，株価の市況コメントでは，「上がる」「下がる」といった価格の単純な特徴だけが表出されるわけではなく，株価の短期的または長期的な値動きが言及されることもあれば，それらを同時に考慮した値動きが言及される点である．2 つ目は，テキストが書かれる時間帯によっても言及される情報は様々であり，例えば「大引け」「前引け」などの配信時刻に依存する表現が見られる点である．3 つ目は，市況コメントでは「東証終値 505 円高の 26,476 円」のように， 当日の終値などの価格（26,476 円）がテキスト中で直接言及されることもあれば，前日からの増減幅（505 円）やそれらを切り上げ・切り捨てした価格などが言及される点である．このような数値情報を正確に言及するための方法として，

コピー機構 (See et al., 2017) などを用いて入力の時系列株価データから価格をコピーすることが考えられる．しかし，前日からの増減幅（505 円）のように，過去履歴からの差分を算出するなどの演算操作によって導出された数値情報は，入力データには含まれないためコピー機構では生成することができない．

本研究では，エンコーダ・デコーダモデルを用いて，上記のような多様な特徴を自動抽出してテキスト化するためのエンコード/デコード手法を探求する．まず，株価の短期的・長期的な変化を捉えるために，エンコーダへの入力として短期的および長期的な時系列株価データを与える．デコード時には，テキストが書かれる時間帯に依存する表現を生成するために，時間帯情報を導入する．また，デコーダが数値に言及する際，数値の演算操作を推定して計算することで株価の数値表現を生成する．これにより，市況コメントで見られる株価の増減幅などの演算操作によって導出される数値についても言及することが可能となった．実験では，自動評価および情報性・流暢性に関する人手評価を行い， 提案手法によって上記の特徴を捉えた質の高い株価の市況コメントの生成が可能になることを示した．

1.2.2　数値気象予報からの天気予報コメントの自動生成

本研究では，天気予報コメントの自動生成タスクに取り組む．一般的に天気予報コメントは，気象予報モデルを大規模コンピュータでシミュレーションして得られる数値気象予報を基に人手で記述されている．そこで本研究では，こうした実際の天気予報コメントの制作過程に着想を得て，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ことを目指す．

本研究では，天気予報コメントにおける 3 つの特徴に着目する．1 つ目は，天気予報コメントを記述するためには，数値気象予報のシミュレーション結果に含まれる気圧や降水量，雲量などの複数の物理量とその時間変化を考慮しなければならない点である．2 つ目は，天気予報コメントは一日に複数回にわたって様々なエリアに向けて配信されており，各コメントで言及される情報はその配信時刻や対象エリアに依存する点である．例えば，海辺のエリアを対象とした天気予報コメントでは，波の高さなどについて言及されることがある．3 つ目は，天気予報サイトのユーザは天気予報コメントの情報の有用性を重要視している点である．特に「晴れ」「雨」「曇り」「雪」といった情報はユーザにとって重要であり，適切かつ明示的に言及することが求められる．このような情報へ適切に言及する方法としては，入力データ中の重要な情報を予測するための内容選択モデルを導入することが考えられる．しかし，これまでの研究で広く用いられる内容選択モデルは，入出力データの単語一致に基づいて教師データを作成し，入力データ中の単語を言及すべきか否かという分類問題により学習されるため，データの特徴が異なる数値データを対象とした本研究へ適用することが難しい．

本研究では，エンコーダ・デコーダモデルを用いて，上記の特徴を捉えた上でテキスト化するためのエンコード/デコード手法を探求する．まず，数値気象予報のシミュレーション結果に含まれる様々な物理量やその時間変化を捉えるために，適切なエンコード手法を比較検討する．次に，天気予報コメントの対象エリアや配信時刻に依存する情報について言及するために，対象エリアや配

信時刻といった各コメントのメタ情報をモデルへ与える．また，「晴れ」「雨」「曇り」「雪」といった気象情報をユーザにとって重要な情報と定義し，入力の気象データからこれらの気象情報を予測する内容選択モデルを導入する．本研究では，入力データに依存せず，手がかり語に基づいて出力テキストで言及される気象情報を同定し，それらを教師データとして内容選択モデルを構築する．これにより，単語の表層に基づいて入出力データの対応が取れない場合であっても，入力データから気象情報を予測する内容選択モデルを導入することが可能となった．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．

1.3 本論文の構成

本論文の構成について説明する．2 章では，まず Data-to-Text 生成課題における研究動向について述べる．その後，本研究で取り組む株価の市況コメント生成や天気予報コメント生成の既存研究について説明する．3 章では，1 つ目の研究である時系列株価データからの市況コメントの自動生成について説明する．この章では，まず株価の市況コメントにおける 3 つの特徴を，具体例を挙げて説明する．その後，提案手法として，これらの特徴を捉えてテキスト化するためのエンコーダ・デコーダモデルについて解説する．実験では，ベースラインモデルとの比較により，各特徴の観点において提案手法が優れていることを示す．4 章では，2 つ目の研究である数値気象予報からの天気予報コメントの自動生成について説明する．この章では，まず天気予報コメントにおける 3 つの特徴について説明した後，本研究で使用する気象データの概要を解説する．次に各特徴を考慮するための提案手法を説明する．実験では，ベースラインモデルと比較して，提案手法により性能が向上することを示す．5 章では，まず本論文のまとめを行い，各研究における今後の課題を述べる．そして最後に，2 つの研究を通して気づきを得られた Data-to-Text 生成課題全体における今後の研究の方向性について議論する．

第 2 章

関連研究

2.1 Data-to-Text 生成課題の研究動向

自然言語生成とは，何らかの「データ」を入力として受け取り，その入力に基づいて自然言語を生成する課題である．自然言語生成技術の応用先は非常に広く，例えば，ある言語の文章を異なる言語へ翻訳する機械翻訳 (Sutskever et al., 2014)，ニュース記事等の文書を簡潔にまとめる文書要約 (Narayan et al., 2018)，ある画像の内容を説明する文章を生成する画像キャプション生成 (Vinyals et al., 2015) などの研究課題がこれまで取り組まれている．本研究で取り組む Data-to-Text 生成課題は自然言語生成の一種であり，画像や表，数値等の「非言語データ」を入力として受け取る課題のことを総称して Data-to-Text 生成課題と呼ばれている．これに対して，機械翻訳や文書要約のように，ある言語の文章やニュース記事等の「言語データ」を入力として受け取る課題は Text-to-Text 生成課題と呼ばれている (Gatt et al., 2018).

数値データや構造化データの概要を人間にとって分かりやすく伝えるために，様々なドメインにおいて Data-to-Text 生成技術の研究が行われている (Gatt et al., 2018; Sharma et al., 2022). 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 (Puduppully et al., 2019; Wiseman et al., 2017)，医師や看護師の意思決定補助を目的に臨床検査データから概況テキストを生成する研究 (Banaee et al., 2013b; Portet et al., 2009)，一定期間毎の学習態度を記録した時系列データから学生へのフィードバックのテキストを自動生成する研究 (Gkatzia et al., 2014)，レストランの特徴が記述された表データから紹介文を生成する研究 (Novikova et al., 2017) などが行われている．本研究で取り組む時系列株価データからの市況コメント生成および数値気象予報からの天気予報コメント生成は，それぞれ時系列数値データを対象とした Data-to-Text 生成課題の一種である．

一般的に Data-to-Text 生成技術は，(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 et al., 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, S. Sripada, 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 (Barzilay et al., 2005; Pablo A. Duboue et al., 2001; Pablo Ariel Duboue et al., 2003) 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 (Iso et al., 2019; Liu et al., 2018)．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; Cho et al., 2014; Sutskever et al., 2014) を用いた手法に注目が集まっており，大規模データを用いて学習することで，これまで必要とされていた複雑なルールの記述や特徴量の設計，テンプレートの作成等をすることなく，人間が書いたテキストと遜色ないほど流暢なテキストの生成が可能であることが様々な研究で報告されている (Lebret et al., 2016; Mei et al., 2016b; Sha et al., 2018a)．また，エンコーダ・デコーダモデルは，入力データを受け取るエンコーダを適切に設計することで，画像やテキスト，音声等の様々なデータを扱えるという高い汎用性を備えており，こうした点も様々な形式のデータを扱う Data-to-Text 生成課題において今日の標準的手法になっている理由の一つとして考えられる．

しかしその一方で，エンコーダ・デコーダモデルは生成時の制御が難しく，しばしば入力データの事実と異なる内容や重要性の低い内容のテキストが生成される問題が指摘されている (Tian et al., 2019; Z. Wang et al., 2020; Wiseman et al., 2017)．この問題の要因として W. Li et al., 2022 らは，データとモデルの課題をそれぞれ指摘している．まず前者については，入力データと出力テキストの整合性の欠如が挙げられる．つまり，入力データに忠実でない情報が出力テキストに含まれている課題である．例えば，特定の人物のプロフィールに関する表データから人物紹介文を生成するための WikiBio データセット (Lebret et al., 2016) では，入力の表データに忠実でない人物紹介文が 62% の事例に含まれていると指摘されている (Dhingra et al., 2019)．このような整合性を欠くデータセットでモデル学習することで，結果的に生成モデルの忠実性を低下させる要因となっている (Filippova, 2020)．次に後者のモデルの課題については，主に入力データに対するモデルの表現力の不足が挙げられる．特にエンコーダ・デコーダモデルにおいて，エンコーダは入力データを理解し，それを意味のある表現にエンコードする重要な役割を担っている．しかし，エンコーダの理解能力や表現力が不十分な場合，入力データ中の重要な情報を適切に捉えることができず，結果的に重要でない情報や入力データに対して忠実ではないテキストが生成される要因となっている (Parikh et al., 2020)．その他にも主にデコーダ，すなわちニューラル言語モデルの性質に起因する問題として，Out-of-vocabulary 問題 (Luong et al., 2015) や Exposure bias (C. Wang et al., 2020) 等の問題も指摘されている (W. Li et al., 2022)．Out-of-vocabulary 問題とは，デコーダ側の単語辞書に含まれない単語は生成できないという問題である．ニューラル言語モデルの性質上，モデルで扱える語彙数には上限があるため，学習データで出現頻度の高い単語を優先して単語辞書に登録することが一般的である．すわなち，学習データ中の低頻度語や評価データで初めて出現する単語は辞書に登録されず，本来は異なる単語にもかかわらず画一的に <unk> などの特殊トークンとして扱われる．こうしたデコーダの性質により，結果的に入力データとは異なる内容（単語）がテキスト中に表出され，テキストの正確性の低下に繋がっている．

上記のデータおよびモデルの課題はこれまで多くの研究で取り組まれている．まず前者のデータの課題に対しては，データセット自体の品質向上の取り組みがある (Dušek et al., 2019)．例えば，Nie et al., 2019 らは入力データに対する出力テキストの忠実性を向上するために，入力データ自体を修正する取り組みを行っている．また，Parikh et al., 2020 らは Wikipedia の表データとその説明文のペアからなる ToTTo データセットの構築において，表データに含まれない情報を説明文から取り除くために人手による説明文の修正を行っている．しかしながら，Parikh et al., 2020 らの実験では，ToTTo のような高品質なデータセットを用いた場合であっても，入力データに対して忠実ではないテキストが生成されることが指摘されている．これは後者の入力データに対するモデルの表現力の課題に起因することが示唆される．そのため，後者のモデルの課題に対しては，モデル自体を高度化する取り組みが行われている (Gatt et al., 2018; Sharma et al., 2022)．その一つの方向性として，エンコーダ・デコーダモデルに対して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動きが盛んになってきている (Iso et al., 2019; Lebret et al., 2016; Puduppully et al., 2019; Sha et al., 2018b; Z. Wang et al., 2020)．例えば Ma et al., 2019 らは，表データから説明テキストを生成する Table-to-Text タスクにおける内容選択として，表データ中の言及すべき特定のセルを予測する分類器，すなわち内容選択モデルを導入している．Ma et al., 2019 らの実験では，内容選択モデルによって表データ中の言及すべき重要箇所を明示的に考慮できるようになったことで，生成テキストの品質が大幅に改善されることが報告されている．また，Iso et al., 2019 らは内容プランニングについて，特に情報の取捨選択が必要なデータから長い文書を生成する場合に導入することで，より入力データに忠実なテキスト生成が実現できることを報告している．さらに Wiseman et al., 2017 らは，デコーダ側の課題となっていた Out-of-vocabulary 問題に対応するために，表データに含まれる固有名詞や数詞といった単語を直接参照してテキスト化するためのコピー機構 (See et al., 2017) を導入し，入力データに対する出力テキストの忠実性の向上に取り組んでいる．

このように，これまでの Data-to-Text における生成テキストの正確性の課題に対する取り組みでは，モデル自体を高度化するという研究の有用性が数多く示されてきた．その代表的な例が Table-to-Text タスクを中心に広く用いられる内容選択モデル (Ma et al., 2019) やコピー機構 (See et al., 2017) である．しかし，これらの手法は数値や画像といった表データとは異なる形式のデータに対して適用する際に課題に直面する．なぜならこれらの手法は，表データとテキストのように入力データと出力テキストの単語の表層に基づいて対応が取れることを前提としているためである．例えば，Table-to-Text タスクにおける内容選択モデルは，表データとテキストの単語一致に基づいて事前に教師データを作成し，表データ中の内容（単語）を言及すべきか否かといった分類問題により学習されている (Ma et al., 2019)．しかし，表データと異なる特徴を持つ数値データを扱う場合，入力データとテキストの対応関係を直接的に獲得することは困難であるため，前述の方法によって内容選択モデルの教師データを作成することができない．そのため本研究では，入力データに依存せずに手がかり語に基づいて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方法を提案する．また，コピー機構は入力データに含まれる単語や数値を直接コピーしてテキスト化するために広く用いられているが，時系列

数値データに対する説明テキストでは，株価の市況コメントのように入力データ（時系列株価データ）に含まれる株価の終値等を直接的に言及することもあれば，株価の上げ幅や下げ幅のように入力データを演算した値が言及されることもある．コピー機構は，前者のように入力データに含まれる数値は生成できるものの，後者のように演算が必要な数値は生成することができない．そのため本研究では，生成モデルが数値に言及する際に入力データから適切な数値を導出するための演算操作を推定し，それを計算することでより正確に数値に言及する方法を提案する．

2.2 株価の市況コメントの自動生成

次に 1 つ目の研究である株価の市況コメントの生成に関連する既存研究について説明する．

Kukich, 1983 らは，株価の値動きや過去の価格との比較，数値表現などの特徴を持った市況コメントの自動生成に取り組む研究の先駆けとして， 日足の株価データベースから市況コメントを生成するための複数のルールを組み合わせた手法を提案した．具体的には， まず，入力である日足の株価データベースの中から 120 個のルールに基づいて言及するべき値動きや価格等の内容選択を行い， 次に 16 個のルールに基づいて内容プランニングを行う．そして，選択した内容および 109 個のルールとフレーズ辞書に基づいて，使用するフレーズや述語句の統語形式，主語の照応関係等の選択を行うことで市況コメントを生成する．このように，Kukich, 1983 らの研究では，人手で数多くのルールを記述する必要がある．一方，我々が提案するエンコーダ・デコーダモデルを用いた手法では，実際の時系列株価データと人手で書かれた市況コメントのペアデータから対応関係を学習し，これらに基づいて， 言及するべき値動きや価格等の内容選択および使用するフレーズ (単語列) の選択等を行うため，数多くのルールは不要である．

同様に，K. Aoki and Kobayashi, 2016 らは，日経平均株価を対象に，株価の値動きを説明するテキストの生成に取り組んでいる．K. Aoki and Kobayashi, 2016 らは本研究と同様に，時系列株価データと市況コメントのペアデータから学習した対応関係を基にテキスト生成を行う機械学習ベースの手法を提案している．具体的には，クラスタリング手法により算出した時系列株価データの類似度を基に，重み付けされた bi-gram 言語モデルを生成し，その言語モデルを用いてテキスト自動生成を行う．また，K. Aoki and Kobayashi, 2016 らの研究では，時系列株価データの類似度に基づいて言及する値動きを選択する内容選択タスクおよび言語モデルによる表層化の 2 つを個々に取り組んでいる．一方，本研究ではエンコーダ・デコーダモデルによりこれらのサブタスクを同時に取り組むことが可能となる．加えて，K. Aoki and Kobayashi, 2016 らの研究では，市況コメントにおいて言及される株価の終値や値上げ幅といった数値への言及を行う取り組みが行われていない．これに対し本研究では，数値の変動を概況するだけでなく，入力の時系列数値データを参照した上で，実際の数値へ言及を行うテキストを生成する手法を提案する．

また，T. Aoki et al., 2018 らは，市況コメントにおいてしばしば言及される株価の変化要因の生成に取り組んでいる．ここで変化要因とは，「日経平均、反落で始まる　下げ幅 100 円超、欧米株安・円上昇で」のように，市況コメントの主な記述対象である株価データ (日経平均株価) の値動きに影響したとされる外国株式や原油価格などの情報のことを指している．K. Aoki,

Miyazawa, et al., 2019 らは，時系列株価データに加えて，市況コメントの生成内容を表すトピックを入力として与えることで，市況コメント生成タスクにおける生成文の内容制御に取り組んでいる．さらに，D. Zhang et al., 2018 らは，人手で書かれた文章のように自然で多様なテキストの生成を目的に，株価の値動きの方向 (値上がり，値下がり) やその変動幅を表すための動詞を適切に選択するための研究に取り組んでいる．これらの研究では，市況コメント生成タスクにおける，変化要因の記述 (T. Aoki et al., 2018) や生成文の内容制御 (K. Aoki, Miyazawa, et al., 2019)，多様な表現を用いたテキスト生成 (D. Zhang et al., 2018) を行うことを目的としている．一方，本研究では市況コメントにおける価格の履歴を参照する表現や時間帯に依存する表現，株価の数値表現などの様々な特性を表出するテキストの生成に取り組んでおり，これらの研究とは目的が異なる．

2.3 天気予報コメントの自動生成

次に 2 つの目の研究である天気予報コメントの生成に関連する既存研究について説明する．

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, S. Sripada, 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 (Angeli et al., 2010; Belz, 2007; Mei et al., 2016b)．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 2.1 の SUMTIME-METEO (S. Sripada, Reiter, Hunter, et al., 2003) や図 2.2 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 (S. G. Sripada et al., 2002) において作成された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[*1]．本データの特徴として，図 2.1 のように，海上の風や波の高さに関

[*1] 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

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 2.1: SUMTIME-METEO の例

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[*2]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 2.2 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

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.

[*2] https://www.weather.gov/

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 2.2: WEATHERGOV の例

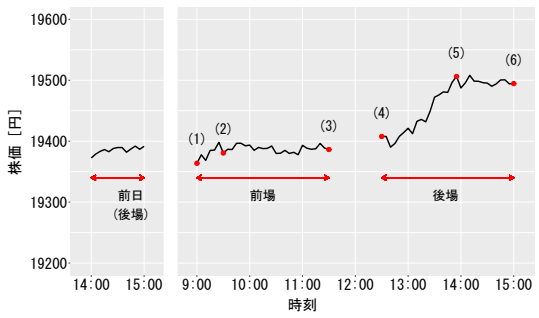
いう実際の制作過程を模したタスク設定となっている．このように，数値気象予報のシミュレーション結果や気象観測値等の生データを入力として考えることで，天気予報コメントの自動生成タスクにおいて次の 2 つの利点があると考えられる．まず，1 つ目に，SUMTIME-METEO や WeatherGov ではデータを構造化することで情報量の劣化が懸念されるが，本研究で使用する数値気象予報のシミュレーション結果や気象観測値等の生データは情報量が多く，これら全てを活用することでより正確な天気予報コメントの生成が期待できる．2 つ目に，数値気象予報のシミュレーション結果から人手またはシステムを介して事前に構造化データを作成するという手間がないため，天気予報コメントの制作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．その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制作作業の自動化において障壁になることが懸念される．したがって本研究では，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．

第 3 章

時系列株価データからの市況コメントの自動生成

3.1　研究概要

本研究では，日経平均株価の市況コメントを生成するタスクを例として，時系列数値データから多様な特徴を抽出し，データの概要をテキスト化する手法を提案する．本研究では，日経平均株価の市況コメントの自動生成を，時系列株価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，機械翻訳や文書要約などの系列生成タスクで広く用いられているエンコーダ・デコーダモデル (Sutskever et al., 2014) を使用する．



	時刻	市況コメント
(1)	09:00	日経平均、続落で始まる
(2)	09:29	日経平均、上げに転じる
(3)	11:30	日経平均、続落 前引けは 5 円安の 19,386 円
(4)	12:30	日経平均、午後は上昇で始まる
(5)	13:54	日経平均、上げ幅 100 円超える
(6)	15:00	日経平均、反発 大引けは 102 円高の 19,494 円

図 3.1: 日経平均株価と市況コメント

図 3.1 に日経平均株価の時系列株価データと市況コメントの例を示す．この例のように，株価の市況コメントなどの時系列数値データの概況テキストでは，「上がる」，「下がる」といった単純な特徴だけが表出されるわけではない．過去のデータの履歴や，テキストが書かれる時間帯によって言及すべき内容は様々である．また，数値の時系列データの場合，時系列中の数値や，過去との差分を計算した値が言及されることが多々ある．例えば，図 3.1 の市況コメントでは，「続落」，「反発」のように価格の履歴を参照する表現 (1，3，6)，「上げに転じる」のように時系列データの変化

を示す表現 (2),「始まる」,「前引け」,「午後」,「大引け」などテキストが書かれる時間帯に依存する表現 (1，3，4，6) が見られる．また，数値に言及する場合は，価格が直接言及される (3，6) こともあれば，履歴からの差分 (3，6) や，切り上げ・切り捨てした値 (5) が用いられることもある．

本研究では，株価の市況コメントにおけるこれらの特性を踏まえ，データから多様な特徴を自動抽出し，テキスト化するためのエンコード/デコード手法を提案する．まず,「続落」,「上げに転じる」といった時系列株価データの過去の履歴や変化を捉えるために，株価の短期的および長期的な時系列データを使用する．次に,「前引け」,「大引け」といった市況コメントが記述される時間帯に依存する表現を生成するために，デコード時に時刻情報を導入する．加えて,「19,386 円」,「100 円」といった株価の終値や前日からの変動幅などの数値を市況コメントで言及するために，入力である時系列株価データ中から適切な数値を出力するための演算操作を推定し，計算することで数値の出力を行う．

実験では，日経平均株価の時系列株価データと人手で書かれた市況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，実際の市況コメントと生成テキストの一致度合いを評価するための BLEU，および,「続落」,「前引け」などの表現を正しく出力でき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大幅に性能が向上することを確認した．さらに，人手評価では，テキストの流暢性と情報性の観点において，提案手法により株価の市況コメントにおける上記のような多様な特徴を捉えた質の高いテキストを生成できることを示した．

3.2 提案手法

近年，機械翻訳 (Cho et al., 2014) や文書要約 (Rush et al., 2015) などの様々な系列生成タスクにおいて，エンコーダ・デコーダモデル (Sutskever et al., 2014) を用いた手法が提案され，有用性が示されている．本研究では，時系列株価データに対する市況コメントの生成を，時系列株価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，エンコーダ・デコーダモデルを用いた手法を提案する．本研究では，エンコーダとして一般的に利用されている多層パーセプトロン (Multi-Layer Perceptron: MLP)，畳み込みニューラルネットワーク (Convolutional Neural Network: CNN)，リカレントニューラルネットワーク (Recurrent Neural Network: RNN) のうちいずれかを採用し，それぞれの性能の比較を行う．また，デコーダには，テキスト生成タスクにおいて広く使われているリカレントニューラルネットワーク言語モデル (Recurrent Neural Network Language Model: RNNLM) を利用する．

時系列株価データの市況コメントを記述する際には，モデルでは，時系列データの絶対的・相対的な変化や最大値・最小値といった特徴を，異なるタイムスケールで捉える必要がある．また，市況コメントでは,「前引け」,「大引け」などのテキストが書かれる時間に依存する表現が用いられることや，株価の終値や変動幅などの数値について言及されることがある．本研究では，このような時系列株価データの多様な特徴を自動抽出してテキスト化するために，標準的なエンコーダ・デコーダモデルに対して 3 つの手法を提案する．

本研究で提案するモデルの概要を図 3.2 に示す．提案モデルでは，時系列データの様々な変化を

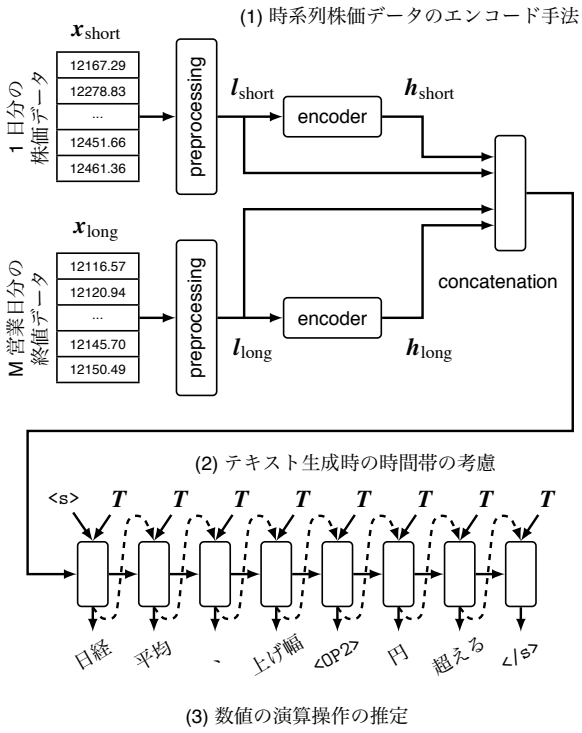


図 3.2: 提案モデルの概要

異なるタイムスケールで捉えるために，短期的な時系列データ $\boldsymbol{x}_{\text{short}}$ および長期的な時系列データ $\boldsymbol{x}_{\text{long}}$ を入力として利用する．また，テキストが書かれる時間に依存する表現を生成するために，デコード時にテキストを記述する時間帯の情報 ($\boldsymbol{T}$) を利用する．加えて，市況コメントの生成時に株価の終値や変動幅などの実際の数値に言及する際には，入力した時系列データ中から正しい数値を出力するための演算操作 (<OP2> 等) を推定し，計算することで数値の出力を行う．

以降では，(1) 時系列株価データのエンコード手法，(2) テキスト生成時の時間帯の考慮，(3) 数値の演算操作の推定について詳細を説明する．

3.2.1 時系列株価データのエンコード手法

本研究では，時系列株価データとして，日経平均株価を使用する．株価の短期的または長期的な数値の変動を捉えるために，短期的な時系列データとして，N タイムステップからなる 1 日分の株価データ $\boldsymbol{x}_{\text{short}} = (x_{\text{short},\,i})_{i=0}^{N-1}$，長期的な株価データとして，過去 M 営業日分の終値 $\boldsymbol{x}_{\text{long}} = (x_{\text{long},\,i})_{i=0}^{M-1}$ を入力として利用する．

画像処理分野 (Cun et al., 1990) や自然言語処理分野 (Vijayarani et al., 2015) などの様々な分野において，機械学習モデルの汎化性能やデータに含まれるノイズを除去するために，データに対して前処理を行うことが一般的である (Banaee et al., 2013a; G. P. Zhang et al., 2005)．本研究でも同様に，数値データである時系列株価データに対して，前処理を行う．

数値データの前処理手法として，標準化 (*standardization*) と前日との差分 (*moving reference*) (Freitas et al., 2009) を使用する．使用する前処理手法の式を以下に示す:

$$x_i^{\mathrm{std}} = \frac{x_i - \mu}{\sigma}, \tag{3.1}$$

$$x_i^{\mathrm{move}} = x_i - r_i, \tag{3.2}$$

ここで，x_i は数値データである株価を表す．式 (3.1) では，学習に使用する全株価データ $\boldsymbol{x}$ の平均値 μ，標準偏差 σ を用いて標準化を行う．式 (3.2) では，前日の終値からの価格の変動を捉えるために，前日の終値 r_i から各タイムステップの価格 x_i の差分を計算する．

次に，時系列株価データの前処理とエンコードの手順について説明する．まず，前処理は，1 日分の株価データ $\boldsymbol{x}_{\mathrm{short}}$，$M$ 営業日分の終値データ $\boldsymbol{x}_{\mathrm{long}}$ に対して行い，数値ベクトル $\boldsymbol{l}_{\mathrm{short}}$，$\boldsymbol{l}_{\mathrm{long}}$ をそれぞれ作成する．次に，作成した数値ベクトルをそれぞれエンコーダへ入力し，エンコーダの出力状態ベクトル $\boldsymbol{h}_{\mathrm{short}}$，$\boldsymbol{h}_{\mathrm{long}}$ を獲得する．続いて，前処理により作成した数値ベクトルとエンコーダの出力状態ベクトルを結合し，*multi-level representation* ベクトル (Mei et al., 2016a) を作成する．この multi-level representation について，Mei et al., 2016a は，入力データの高レベルな表現 ($\boldsymbol{h}_{\mathrm{short}}$，$\boldsymbol{h}_{\mathrm{long}}$) と低レベルな表現 ($\boldsymbol{l}_{\mathrm{short}}$，$\boldsymbol{l}_{\mathrm{long}}$) を同時にモデルで考慮することで，モデルが入力データ中の重要な情報を選択する性能が向上したことを報告している．株価の市況コメントの記述においても，入力データである短期的および長期的な時系列株価データ中の 重要な値動きを捉えることが必要とされる．そのため，本研究でも同様に，multi-level representation を採用し，デコーダである RNNLM の初期状態 $\boldsymbol{s}_0$ を次のように設定する:

$$\boldsymbol{s}_0 = \boldsymbol{l}_{\mathrm{short}} \oplus \boldsymbol{l}_{\mathrm{long}} \oplus \boldsymbol{h}_{\mathrm{short}} \oplus \boldsymbol{h}_{\mathrm{long}}. \tag{3.3}$$

ここで，$\oplus$ は連結演算子を表している．

2 つの前処理手法を用いる場合，短期的及び長期的な時系列株価データ $\boldsymbol{x}_{\mathrm{short}}$， $\boldsymbol{x}_{\mathrm{long}}$ に対してそれぞれの前処理を適用し，4 つの数値ベクトル $\boldsymbol{l}_{\mathrm{short}}^{\mathrm{move}}$，$\boldsymbol{l}_{\mathrm{short}}^{\mathrm{std}}$, $\boldsymbol{l}_{\mathrm{long}}^{\mathrm{move}}$，$\boldsymbol{l}_{\mathrm{long}}^{\mathrm{std}}$. を作成する．次に，それぞれの数値ベクトルに対して独立のエンコーダを使用し，計 4 つのエンコーダの出力状態ベクトルを獲得する．この時，デコーダである RNNLM の初期状態 $\boldsymbol{s}_0$ は次のように設定する:

$$\boldsymbol{s}_0 = \boldsymbol{l}_{\mathrm{short}}^{\mathrm{move}} \oplus \boldsymbol{l}_{\mathrm{short}}^{\mathrm{std}} \oplus \boldsymbol{l}_{\mathrm{long}}^{\mathrm{move}} \oplus \boldsymbol{l}_{\mathrm{long}}^{\mathrm{std}} \oplus \boldsymbol{h}_{\mathrm{short}}^{\mathrm{move}} \oplus \boldsymbol{h}_{\mathrm{short}}^{\mathrm{std}} \oplus \boldsymbol{h}_{\mathrm{long}}^{\mathrm{move}} \oplus \boldsymbol{h}_{\mathrm{long}}^{\mathrm{std}}. \tag{3.4}$$

また，時系列株価データのような時系列数値データから，数値の変動等の特徴を抽出するためのエンコード手法として，いくつかの方法が考えられる．本研究ではエンコーダとして，MLP, CNN，RNN のいずれかを用いる．本研究では，実験において，時系列数値データの特徴抽出手法として有用なエンコード手法を比較検討する．

3.2.2 テキスト生成時の時間帯の考慮

時系列データの概況テキストは，テキストが書かれる時間帯に依って言及すべき内容は様々である．例えば株価の市況コメントの場合，一般的に，図 3.1 中の (1)，(6) のように，取引が始まる時

間帯には「前日から価格がどのように変動したか」，取引が終了する時間帯には「値上げ幅と終値はいくらか」等が言及される．

J. Li et al., 2016 ら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は，デコード時にペルソナ情報を追加的に入力することで，指定したペルソナの特徴を捉えた単語列を生成できることが報告されている．これらを踏まえ本研究では，デコード時の各タイムステップの状態 $\boldsymbol{s}_j$ に時間帯情報 $\boldsymbol{T}$ の付与を行い，時間帯を考慮したテキストの生成を行う．具体的には，市況コメントが配信される時間帯 (9 時，15 時等) を入力として時間帯情報埋め込みベクトル $\boldsymbol{T}$ を作成し，デコード時の各タイムステップの隠れ状態ベクトル $\boldsymbol{s}_j$ に時間帯情報埋め込みベクトル $\boldsymbol{T}$ を加算する[*1]．

3.2.3 数値の演算操作の推定

RNNLM などの言語モデルを用いたテキスト生成において，一定の出現頻度よりも少ない単語は未知語 (out-of-vocabulary: OOV) として，<unk> などの特殊トークンに置き換えられることが一般的である (Sutskever et al., 2014)．特に，固有名詞や数値などのバリエーションが多い単語は出現頻度が少なくなる傾向があるため，OOV として扱われてしまうことがある．また，これらの単語が OOV として扱われなかった場合であっても，その単語と類似した別の単語を誤って生成してしまうことがある．機械翻訳の分野では，OOV 問題の対策として，出現頻度が少なくなりやすい固有名詞などを入力テキストからコピーを行い，これらの単語を出力するための機構が提案されている (Gulcehre et al., 2016; Luong et al., 2015; See et al., 2017)．

入力の数値データに言及するテキストでは，図 3.1 中の (3，6) のように，入力データに含まれる数値について直接言及することが多い．しかし，それだけではなく，履歴からの差分 (3，6) や，切り上げ・切り捨てした値 (5) が用いられることもある．そのため，入力データから数値をコピーするだけでなく，「差分の計算」等の数値の演算操作が必要となる．しかしながら，通常のモデルでは，このような演算操作を必要とする数値を直接的に生成することができない．そこで本研究では，演算した数値を間接的に生成するために，12 種類の演算トークンを導入する．具体的には，モデルにおいて株価の数値箇所を直接的に予測する代わりに，12 種類の演算トークンのいずれかを推定し，予め定義した各演算トークンに対応する演算操作のルールに基づいて価格の計算を行い，計算結果の価格で演算トークンを置換する．

本手法では，前処理として，学習データのテキスト中の価格箇所を <OP1> や <OP2> 等の演算トークンに置換する．使用する演算トークンは，言及する価格の性質に依って異なる．表 3.1 に事前に定義した演算トークンと対応する演算操作の内容を示す．ここで，次の市況コメントを例として，学習データのテキストの前処理方法について説明する．

(1)　日経平均、反発 午前終値は 227 円高の 16,610 円

[*1] J. Li et al., 2016 の研究では埋め込みベクトルを連結しているが，予備検証において加算の場合でも連結の場合と同様に出力が変化することが明らかになったため，今回は加算を採用した．

表 3.1: 定義した演算トークンと演算操作

演算トークン	操作内容
<OP1>	$x_{\mathrm{long},\,M-1}$ と $x_{\mathrm{short},\,N-1}$ の差を返す
<OP2>	$x_{\mathrm{long},\,M-1}$ と $x_{\mathrm{short},\,N-1}$ の差を 10 の位で切り捨て
<OP3>	$x_{\mathrm{long},\,M-1}$ と $x_{\mathrm{short},\,N-1}$ の差を 100 の位で切り捨て
<OP4>	$x_{\mathrm{long},\,M-1}$ と $x_{\mathrm{short},\,N-1}$ の差を 10 の位で切り上げ
<OP5>	$x_{\mathrm{long},\,M-1}$ と $x_{\mathrm{short},\,N-1}$ の差を 100 の位で切り上げ
<OP6>	$x_{\mathrm{short},\,N-1}$ を返す
<OP7>	$x_{\mathrm{short},\,N-1}$ を 100 の位で切り捨て
<OP8>	$x_{\mathrm{short},\,N-1}$ を 1,000 の位で切り捨て
<OP9>	$x_{\mathrm{short},\,N-1}$ を 10,000 の位で切り捨て
<OP10>	$x_{\mathrm{short},\,N-1}$ を 100 の位で切り上げ
<OP11>	$x_{\mathrm{short},\,N-1}$ を 1,000 の位で切り上げ
<OP12>	$x_{\mathrm{short},\,N-1}$ を 10,000 の位で切り上げ

前処理では，まず始めに，表 3.1 中の全ての演算操作 (12 種類) を行い，各演算トークンに対応する数値を計算する．例えば，ここで，前日の終値 $x_{\mathrm{long},\,M-1}$ を「16,383」，最後のタイムステップの価格 $x_{\mathrm{short},\,N-1}$ を「16,610」とした場合，演算トークン <OP1> に対応する数値は「227」となる．次に，各演算トークンに対応する数値「227」等とテキスト (1) 中の数値「227」,「16,610」を比較し，正解の数値「227」,「16,610」のそれぞれに最も近い数値を計算した演算操作およびその演算トークンを求める．最後に，導出した演算トークンを，正解の数値「227」,「16,610」に対する最適な演算トークンとして見做し，正解の数値と演算トークンを置換することで，以下の前処理済みテキスト (2) を獲得する．

(2)　日経平均、反発 午前終値は <OP1> 円高の <OP6> 円

上記の例では，テキスト (1) 中の「227」は，前日の終値 $x_{\mathrm{long},\,M-1}$ である「16,383」と最後のタイムステップの価格 $x_{\mathrm{short},\,N-1}$ である「16,610」の差を表すため，演算トークン <OP1> に置換し，「16,610」は，最後のタイムステップの価格 $x_{\mathrm{short},\,N-1}$ である「16,610」を表すため，演算トークン <OP6> に置換している．

次に，テスト時における数値の導出方法について説明する．入力として，前日の終値 $x_{\mathrm{long},\,M-1}$ が 14,612 円で最後のタイムステップの価格 $x_{\mathrm{short},\,N-1}$ が 14,508 円の時系列株価データをモデルへ与え，テキスト (3) を生成した場合を考える．

(3)　日経平均、反落で始まる 下げ幅 <OP2> 円超、<OP7> 円台

まず，<OP2> を，前日の終値 $x_{\mathrm{long},\,M-1}$ と最後のタイムステップの価格 $x_{\mathrm{short},\,N-1}$ の差を 10 の位で切り捨てた価格である「100」へ置換する．次に，<OP7> を，最後のタイムステップの価格

$x_{\mathrm{short},\,N-1}$ を 100 の位で切り捨てた価格である「14,500」へ置換する．以上により，テキスト (4) が得られ，これを出力テキストとする．

(4)　日経平均、反落で始まる 下げ幅 100 円超、14,500 円台

このように，株価データと市況コメントのペアデータから対応関係を学習するエンコーダ・デコーダモデルと本課題に対する少量のルールを組み合わせることにより，従来のルールに基づくテキスト生成手法 (Kukich, 1983) と比べて，より少ないルールによって株価の価格や上げ幅等の数値表現を含む市況コメントの生成が可能となる．

3.3　実験設定

3.3.1　データセット

実験には，時系列株価データとして IBI-Square Stocks[*2]から収集した 2013 年 3 月から 2016 年 10 月までの 5 分足の日経平均株価，市況コメントとして日経 QUICK ニュース社が提供する日経平均株価ニュースのヘッドラインテキスト，計 7,351 件を利用した．市況コメントの内，2013 年 3 月から 2016 年 1 月までの市況コメントである 5,880 件を学習データ，2016 年 2 月から同年 10 月までの 730 件，741 件をそれぞれ開発データ，評価データとして利用した．市況コメントの形態素解析には MeCab[*3] (IPA 辞書) を使用し，各形態素を 1 つの語彙とした．また，学習データの市況コメントにおいて，出現回数が 1 回以下の形態素は未知語として扱い，特殊トークン <unk> へ置換を行った．その結果，学習データにおける語彙サイズは 691，平均文長（1 文あたりの平均語数）は 12.5 となった．

人手評価用の評価データとして，上記評価データからランダムに抽出した 100 件の市況コメントおよび株価データを使用した．

実験では，短期的な時系列株価データである 1 日分の株価データのタイムステップ数 N を 62，長期的な時系列株価データである過去 M 営業日分の終値のタイムステップ数 M を 7 とした．本研究では，時系列株価データとして 5 分足の日経平均株価を用いている． 従って，市況コメントが書かれる直近の時間帯から 62 タイムステップ前までの株価を 1 日分の時系列株価データとして設定している．

学習時には，全 62 タイムステップから成る 1 日分の株価データ $\boldsymbol{x}_{\mathrm{short}}$，全 7 タイムステップから成る過去 7 営業日分の終値データ $\boldsymbol{x}_{\mathrm{long}}$ と市況コメントのペアを使用する．テスト時には，株価データのみを用いて市況コメントを生成する．

[*2] http://www.ibi-square.jp/index.htm
[*3] http://taku910.github.io/mecab/

3.3.2 ハイパーパラメータ

エンコーダ・デコーダモデルの学習において，単語埋め込みベクトルの次元は 128，時間帯情報埋め込みベクトルの次元は 64，エンコーダの隠れ状態の次元は 256 とした．エンコーダに CNN を用いる場合，畳み込み層は 1 層とし，入力チャネル数は 1，出力チャネル数は 16，フィルタサイズは 3，活性化関数には ReLU (Glorot et al., 2011) を使用した．同様に，MLP を用いる場合，隠れ層は 3 層とし，活性化関数には Tanh を用いた．また，RNN を用いる場合，1 層の Long-Short Term Memory (LSTM) (Hochreiter et al., 1997) を使用し，活性化関数には Tanh を用いた．

デコーダには，テキスト生成タスクで広く用いられている RNNLM を用いた．RNN には，LSTM を使用し，レイヤ数は 1 層とした．活性化関数には Tanh を使用した．デコーダの隠れ状態の次元数は，使用する前処理手法の数や multi-level representation の有無によって変化する．例えば，式 (3.4) のように，前処理手法として標準化および前日の差分の 2 手法を用い，さらに multi-level representation を導入する場合，デコーダの隠れ状態は $1,162$ 次元となる[*4]．ここで，数値ベクトル $\boldsymbol{l}^{\mathrm{move}}_{\mathrm{short}}$，$\boldsymbol{l}^{\mathrm{std}}_{\mathrm{short}}$ の次元数は 62，$\boldsymbol{l}^{\mathrm{move}}_{\mathrm{long}}$，$\boldsymbol{l}^{\mathrm{std}}_{\mathrm{long}}$ は 7，エンコーダの出力状態ベクトル $\boldsymbol{h}^{\mathrm{move}}_{\mathrm{short}}$，$\boldsymbol{h}^{\mathrm{std}}_{\mathrm{short}}$，$\boldsymbol{h}^{\mathrm{move}}_{\mathrm{long}}$，$\boldsymbol{h}^{\mathrm{std}}_{\mathrm{long}}$ の次元数は 256 である．

モデルの学習時には，デコーダの各タイムステップの入力として，学習データの市況コメントの単語系列をそのまま用いる Teacher Forcing により学習を行った．また，本研究で使用する市況コメントのデータセットは，機械翻訳等の一般的なテキスト生成タスクと比べてデータ規模が小さいことから，事前学習済み言語モデルなどの事前学習の枠組みを導入することによる性能向上が期待できる (Devlin et al., 2019; Radford et al., 2019; Song et al., 2019)．しかし，本研究で使用する市況コメントデータの予備検証において，事前学習を導入しない場合であっても，一定の学習効果が得られることが明らかになったため，単語埋め込みベクトルやエンコーダ・デコーダモデルにおける重みパラメータの事前学習，および，事前学習済み言語モデルの導入は行っていない．

ミニバッチのサイズは 100，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 et al., 2015) ($\alpha = 0.001$，$\beta_1 = 0.9$，$\beta_2 = 0.999$，$\epsilon = 10^{-8}$) を使用した．学習時のエポック数は 30 に設定した．実験結果では，学習における全 30 エポックの内，開発データに対する BLEU が最も高いエポック時の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．

3.3.3 評価指標

実験では，2 つの自動評価指標と人手評価により生成テキストの評価を行った．1 つ目の自動評価指標として，実際の株価の市況コメントと生成されたテキストの一致度合いを測る目的として

[*4] $62 \times 2 + 7 \times 2 + 256 \times 2 + 256 \times 2 = 1,162$ より，2 つの前処理手法および multi-level representation を導入する場合のデコーダの隠れ層は $1,162$ 次元となる．

表 3.2: F 値による評価に用いた表現および各データにおける出現件数

カテゴリ	表現	各データにおける出現件数		
		学習データ	開発データ	評価データ
株価の変動を説明する表現	続伸	988	115	101
	続落	707	100	117
	反発	889	113	110
	反落	884	104	123
	X 円高の	869	135	135
	Y 円安の	758	126	148
	上げに転じる	172	25	20
	下げに転じる	223	19	24
	上げ幅	774	102	85
	下げ幅	876	111	111
時間帯に依存する表現	始まる	1,341	146	163
	前引け	846	135	161
	大引け	1,327	154	140

BLEU (Papineni et al., 2002) を使用した．BLEU の計測には MTEval toolkit[*5]を使用した．スコアの計算においては，4-gram までを考慮し，大文字・小文字は区別している．また，ブートストラップ・リサンプリング法 (Koehn, 2004) により統計的有意差の検定を行った．有意水準は 5% とした．

2 つ目の自動評価指標として，「続落，反発，上げに転じる」といった短期的・長期的な時系列株価データの変動を説明する表現や「始まる，前引け」などの時間帯に依存する表現を評価データの参照テキストと比較して正しく出力できているかを評価するために，F 値を用いて評価を行った．表 3.2 に F 値による評価で用いる表現を示す．ここで表 3.2 において，各データにおける出現件数は，5,880 件の学習データ，730 件の開発データ，741 件の評価データのそれぞれにおける各表現の出現件数を表している．F 値による評価で用いる表現の選定においては，学習データの市況コメントにおける株価の変動を説明する表現および時間帯に依存する表現のうち，比較的に出現回数が多い 13 種類の表現を選定した．

人手評価では，生成テキストの情報性と流暢性について評価するために，金融工学の専門家の 1 人に評価を依頼した．また，提案手法及び人手により生成した市況コメントの品質の違いを評価するために，システム名を伏せて， 両者の生成テキストおよび対応する短期的・長期的な時系列株価データ $\boldsymbol{x}_{\text{short}}$，$\boldsymbol{x}_{\text{long}}$ を評価者に提示した．人手評価では，情報性と流暢性の 2 つの観点につい

[*5] https://github.com/odashi/mteval

て，0 または 1 の 2 段階のスコア付けを行った．ここで，1 は情報性または流暢性が高いことを表す．情報性の評価では，評価者は生成テキスト及び対応する時系列株価データを参照し評価を行った．具体的には，時系列株価データの重要な値動きや値動きの概況について適切に述べている生成テキストを情報性が高いテキストとして定義した．流暢性の評価では，評価者は生成テキストのみを与え，テキストの可読性の観点で評価を行う．すなわち，株価の値動きについて生成テキストで述べられている内容の正しさにかかわらず評価を行う．また，数値の演算操作の推定を行わない場合，生成テキストにおける数値箇所が未知語 (<unk>) として出力されることで，流暢性の評価に影響する恐れがある．そのため本研究では，生成テキストにおける <unk> には適切な数値や文字列が入っているものとして評価を行うよう評価者に説明した．

また，人手で書かれた市況コメントでは，本研究で使用する時系列株価データには含まれてない情報について記述することがある．例えば，人手で書かれた市況コメントである「日経平均、反落して始まる　米株安や円高で、下げ幅 100 円超える」には，「米株安や円高で」といった外部情報が含まれる．システムへの入力として与える時系列株価データからこのような外部情報を予測することは不可能であることから，人手評価では，生成テキストに含まれる外部情報を無視して評価するように評価者へ依頼した．

3.3.4　比較モデル

表 3.3，3.4 に提案モデルおよび比較モデルの一覧を示す．実験では，まず，エンコード手法の検討として，MLP，CNN，RNN のそれぞれをエンコーダとしたモデル (*mlp-enc*，*cnn-enc*，*rnn-enc*) の比較を行う．

次に，入力の時系列データから短期的及び長期的な変化を捉える能力があるかを確認するために，短期的な株価データ $\boldsymbol{x}_{\text{short}}$ または長期的な終値データ $\boldsymbol{x}_{\text{long}}$ を使用しないモデル (*-short*，*-long*) の比較を行う．また，数値データの表現手法の有用性を確かめるために，*mlp-enc* モデルをベースとして，各前処理手法 (標準化，前日との差分)，multi-level representation を使用しない各モデル (*-std*，*-move*，*-multi*) の評価を行う．

最後に，数値の演算操作の推定手法，および，時間帯情報の入力手法の有用性を確かめるために，各手法を使用しないモデル (*-num*，*-time*) の評価を行う．ベースラインとして，1 日分の株価データのみを入力として，エンコーダに MLP，前処理手法に標準化と前日との差分を使用するモデル (*baseline*) を用いた．

3.4　実験結果

BLEU，F 値による評価の実験結果を表 3.5，表 3.6 にそれぞれ示す．また，各モデルの出力例と人手で書かれた市況コメント (*Human*) を図 3.3 に示す．

表 3.3: 実験で使用した提案モデルの概要

モデル		mlp-enc	cnn-enc	rnn-enc
エンコーダ		MLP	CNN	RNN
入力	$\boldsymbol{x}_{\text{short}}$	✓	✓	✓
	$\boldsymbol{x}_{\text{long}}$	✓	✓	✓
前処理	標準化	✓	✓	✓
	前日との差分	✓	✓	✓
	Multi-level	✓	✓	✓
	演算操作	✓	✓	✓
	時間帯情報	✓	✓	✓

表 3.4: 実験で使用した比較モデルの概要

モデル		baseline	-short	-long	-std	-move	-multi	-num	-time
エンコーダ		MLP	MLP	MLP	MLP	MLP	MLP	MLP	MLP
入力	$\boldsymbol{x}_{\text{short}}$	✓	−	✓	✓	✓	✓	✓	✓
	$\boldsymbol{x}_{\text{long}}$	−	✓	−	✓	✓	✓	✓	✓
前処理	標準化	✓	✓	✓	−	✓	✓	✓	✓
	前日との差分	✓	✓	✓	✓	−	✓	✓	✓
	Multi-level	−	✓	✓	✓	✓	−	✓	✓
	演算操作	−	✓	✓	✓	✓	✓	−	✓
	時間帯情報	−	✓	✓	✓	✓	✓	✓	−

表 3.5: 各モデルの評価データに対する BLEU スコア

モデル	baseline	mlp-enc	cnn-enc	rnn-enc	-short	-long	-std	-move	-multi	-num	-time
BLEU	0.243	**0.464**	0.449	0.454	0.380	0.433	0.455	0.393	0.435	0.318	0.395

3.4.1 時系列株価データのエンコードおよび表現手法の効果

まず，時系列株価データのエンコード手法の検討として，MLP，CNN，RNN のそれぞれをエンコーダとしたモデル (*mlp-enc*，*cnn-enc*，*rnn-enc*) の比較を行う．BLEU による自動評価では，MLP をエンコーダとした提案手法 (*mlp-enc*) がベースラインを含めたその他全てのモデル

表 3.6: 各表現に対する F 値

	baseline	mlp-enc	cnn-enc	rnn-enc	-short	-long	-std	-move	-multi	-num	-time
続落	0.380	0.771	0.780	0.770	0.426	0.250	0.683	0.449	0.726	**0.803**	0.770
続伸	0.405	0.735	0.689	0.697	0.422	0.189	0.670	0.446	0.717	**0.748**	0.636
反発	0.411	0.771	0.755	0.774	0.552	0.512	0.786	0.597	0.782	**0.814**	0.723
反落	0.369	**0.768**	0.743	0.727	0.496	0.554	0.695	0.519	0.623	0.753	0.675
X 円高の	0.584	0.782	0.777	0.769	0.503	**0.789**	0.782	0.615	0.714	0.786	0.633
X 円安の	0.506	0.764	0.770	0.754	0.411	**0.776**	0.750	0.596	0.689	0.747	0.612
上げに転じる	0.125	0.431	**0.500**	0.444	0.000	0.000	0.148	0.000	0.370	0.488	0.286
下げに転じる	0.000	0.353	0.324	**0.400**	0.000	0.333	0.235	0.000	0.254	**0.400**	0.280
上げ幅	0.586	**0.702**	0.667	0.676	0.408	0.663	**0.702**	0.566	0.642	0.693	0.632
下げ幅	0.593	**0.742**	0.669	0.718	0.482	0.665	0.725	0.534	0.674	0.707	0.609
始まる	0.589	0.599	**0.724**	0.696	0.661	0.696	0.648	0.677	0.636	0.693	0.568
前引け	0.629	0.878	**0.885**	0.879	0.881	0.881	0.875	0.866	0.863	0.876	0.587
大引け	0.471	**0.964**	0.960	**0.964**	**0.964**	**0.964**	**0.964**	**0.964**	0.960	**0.964**	0.561

(*baseline*，*cnn-enc*，*rnn-enc* 等) と比較して，有意水準 5% で統計的有意に BLEU スコアが高かった．また，表 3.6 の各表現に対する F 値を比較すると，MLP，CNN，RNN のそれぞれをエンコーダとした 3 モデルにおいて，MLP をエンコーダとした *mlp-enc* がより多くの表現を正しく出力できている．特に，*mlp-enc* は，「日経平均、反発で始まる　上げ幅100 円超」や 「日経平均、一時下げ幅100 円超える」のように，短期的な株価の変動を説明する際に用いられることが多い表現である「上げ幅」，「下げ幅」に対する F 値が他の 2 つのモデル (*cnn-enc*，*rnn-enc*) に対して大きく上回っていた．これらの結果から，単語の一致率を基にした評価指標である BLEU において，MLP をエンコーダとした *mlp-enc* が CNN，RNN のそれぞれをエンコーダとした *cnn-enc*，*rnn-enc* よりも BLEU スコアが向上したことが推察できる．

次に，時系列株価データの表現手法の検討として，数値データの前処理手法として使用した標準化および 前日との差分の比較を行う．表 3.5 の BLEU スコアによる自動評価では，標準化と前日との差分の両方を前処理手法として用いる *mlp-enc* がいずれかの前処理を用いないモデル (*-std*，*-move*) よりも BLEU が高いことが分かった．表 3.6 の F 値による評価においては，両者の前処理手法を用いる *mlp-enc* は「上げに転じる」，「 下げに転じる」等の株価の値動きに言及する表現について，その他の 2 つのモデル (*-move*，*-std*) よりも適切に出力できることが分かった．

また，各前処理手法の有用性を検証するために，両方の前処理手法を用いる *mlp-enc* といずれかの前処理を用いない*-std* および*-move* の BLEU スコアに着目すると，*-move* によって生成した市況コメントは*-std* と比べて，*mlp-enc* よりも大幅に BLEU スコアが低下している．同様に，表 3.6 の F 値による評価において，*-move* は*-std* と比べて，*mlp-enc* よりも「続落」，「反発」等の株価

の変動を説明する表現に対する F 値が大幅に低下している．以上のことから，数値データの前処理手法である前日との差分は，標準化よりも時系列株価データの変動を捉えて株価の値動きを説明する表現を生成する精度に大きく貢献していることが考えられる．

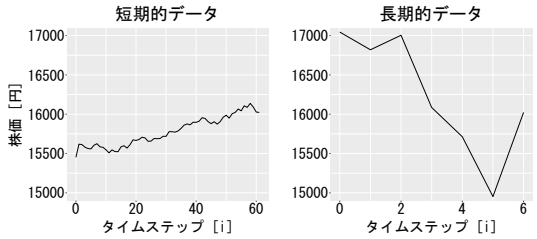
続いて，入力の時系列株価データの表現方法として用いた multi-level representation の効果について比較する．3.2.1 項で述べたとおり，multi-level representation は， エンコーダの出力ベクトルである高レベルな表現 ($\boldsymbol{h}^{\mathrm{move}}_{\mathrm{short}}$，$\boldsymbol{h}^{\mathrm{std}}_{\mathrm{short}}$ 等) と前処理手法により作成した数値ベクトルである低レベルな表現 ($\boldsymbol{l}^{\mathrm{move}}_{\mathrm{short}}$ $\boldsymbol{h}^{\mathrm{std}}_{\mathrm{short}}$ 等) を同時に考慮することで，入力データ中の重要な情報を選択する性能が向上することを期待して導入したベクトル表現手法である．表 3.5 の BLEU スコアによる自動評価では，multi-level representation を用いないモデル (*-multi*) の BLEU スコアが *mlp-enc* よりも低下することを確認した．また，表 3.6 の F 値による評価では，*-multi* は *mlp-enc* と比べて，「続落」，「反落」，「下げに転じる」等の株価の値動きを説明する表現に対する F 値が低下することが分かった．以上のことから，multi-level representation ベクトルを導入することで，モデルが時系列株価データ中の重要な変動を捉える性能が向上し，自動評価指標である BLEU や F 値の改善に寄与したことが考えられる．

さらに，入力データとして短期的及び長期的な時系列株価データを用いることでモデルが株価の様々な変化を異なるタイムスケールで捉えることができているかを検証するために，短期的及び長期的な時系列株価データ $\boldsymbol{x}_{\mathrm{short}}$，$\boldsymbol{x}_{\mathrm{long}}$ を両方用いたモデル (*mlp-enc*，*rnn-enc* 等) とそれぞれ用いないモデル (*-short*，*-long*) の比較を行う．まず，表 3.5 の BLEU による自動評価では，両方のデータを用いた *mlp-enc* は，いずれかを用いない*-short* および*-long* と比べて BLEU スコアが有意に向上することを確認した．特に，短期的な時系列株価データ $\boldsymbol{x}_{\mathrm{short}}$ を入力に与えないモデル (*-short*) では，BLEU スコアが著しく低下した．また，表 3.6 の各表現に対する F 値の評価では，両方の時系列株価データを用いたモデル (*mlp-enc*，*rnn-enc* 等) は，*-short* 及び*-long* と比較して「続落，反発，上げに転じる」等の短期的及び長期的な株価の変化を説明する表現を正しく出力できることが分かった．これらの結果より，入力データとして短期的及び長期的な時系列株価データを与えることで，モデルが株価の様々な変化を捉える性能が向上し，値動きを説明する表現を適切に生成する精度の改善に寄与したことが考えられる．

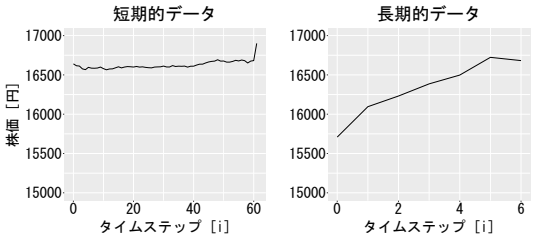
3.4.2 時間帯の考慮手法の効果

次に，モデルが市況コメントが書かれる時間帯に依存する表現を用いて 株価の値動きを適切に説明する性能が向上することを期待して導入した時間帯情報埋め込みベクトル $\boldsymbol{T}$ の効果について検証を行う．具体的には，時間帯情報を用いたモデル (*mlp-enc*) とそれを用いないモデル (*-time*) の比較を行う．

まず，表 3.5 の BLEU による自動評価では，デコード時のデコーダの状態に時間帯情報を付与していない*-time* モデルは，*mlp-enc* と比べて BLEU が有意に低下することを確認した．次に，表 3.6 において，時間帯情報を考慮するモデル (*rnn-enc*，*-num* 等) と時間帯情報を考慮しないモデル (*-time*) を比較すると，*-time* は「始まる，前引け，大引け」といった時間帯に関して言及を行



(a) 2016/2/15 の日経平均株価の値動き



(b) 2016/7/21 の日経平均株価の値動き

モデル	流暢性	情報性	生成テキスト
baseline	1	0	日経平均、反発 前引けは 81 円高の<unk>円
mlp-enc	1	1	日経平均、大幅反発 大引けは 1,069 円高の 16,022 円
Human	1	1	日経平均、大幅反発 大引けは 1,069 円高の 16,022 円

(c) 2016/2/15 15:00 の株価に対する生成テキスト

モデル	流暢性	情報性	生成テキスト
baseline	1	0	日経平均、続伸で始まる 上げ幅 100 円超える
mlp-enc	1	1	日経平均、上げ幅 200 円超える
Human	1	1	日経平均、上げ幅 200 円超す

(d) 2016/7/21 9:00 の株価に対する生成テキスト

図 3.3: 短期的・長期的な時系列株価データおよび各モデルの生成テキスト

う表現を正しく出力できていないことが分かる．

また，表 3.6 より，時間帯情報を考慮しない-*time* は時間帯情報を考慮する *mlp-enc* と比べて，時間帯に依存する表現だけでなく，「X 円高の」，「X 円安の」等の株価の変動幅に言及する表現の予測精度も低下することが分かった．これは，「日経平均、大幅続伸　前引けは251 円高の12,219 円」のように, 株価の市況コメントにおいて，「前引け」，「大引け」は， 前日の終値からの変動幅について言及する際の「X 円高の」，「X 円安の」といった表現とともに用いられる傾向が強いためであると考えられる．つまり，時間帯情報を与えないことで，モデルが「前引け」，「大引け」について言及する性能が低下し，同様に「X 円高の」，「X 円安の」といった表現の予測性能が低下したことが推察できる．

以上のことから，モデルへ時間帯情報を導入することにより，時間帯に依存する表現を適切に用いつつ株価の値動きを説明する性能の改善に貢献することが考えられる．

表 3.7: -num および mlp-enc モデルによる市況コメントの生成例

モデル	生成テキスト
-num	日経平均、続落　前引けは 126 円安の<unk>円
mlp-enc	日経平均、続落　前引けは 56 円安の 17,047 円
Human	日経平均、続落　午前終値は 56 円安の 17,047 円

表 3.8: 数値表現の正解・不正解数

モデル	正解	不正解	合計
mlp-enc	640	223	863
-num	92	771	863

3.4.3　数値の演算操作の推定手法の効果

モデルが生成した株価の市況コメントにおいて，株価の終値や変動幅などの数値を適切に出力できているかを検証するために，数値の演算操作の推定手法を導入したモデル (*mlp-enc*) と導入していないモデル（*-num*）の比較を行う．各モデル (*mlp-enc*，*-num*) による市況コメントの生成例および人手で書かれた参照テキスト (*Human*) を表 3.7 に示す．

まず，表 3.5 の BLEU による自動評価では，演算操作の推定手法を導入していない*-num* は，本手法を導入した *mlp-enc* と比べて BLEU スコアが大きく低下することが分かった．ここで，*-num* と *mlp-enc* の差分は，数値の演算操作の推定手法のみであることに注意されたい (表 3.3，表 3.4)．本手法を導入した *mlp-enc* では，株価等の数値について言及する際，予測した演算操作に基づいて実際の値を計算し出力する．一方，本手法を導入していない*-num* や *baseline* 等では，数値の演算操作の推定を行わず，デコーダである RNN 言語モデルから数値を “単語” として出力する．そのため表 3.7 のように，演算操作の推定を導入していない*-num* によって生成した市況コメントでは，数値として言及すべき箇所において <unk> や適切ではない数値 (126 円) が出力される事例が多く，単語の一致率を基にした BLEU による自動評価において，*mlp-enc* よりも低いスコアとなったことが考えられる．しかし，表 3.6 の株価の値動きを表す表現の F 値によると，*mlp-enc* は *-num* と比べて，「続落」，「反発」等の表現において F 値が劣化することが分かった．この結果から，数値表現を演算トークンへ置き換えることで，これらの表現の予測性能に影響を与えることが推察できる．

次に，数値の演算操作によってどの程度の株価の数値表現を正しく出力できているかについて分析を行う．具体的には，参照テキストに含まれる数値表現のうち，生成テキストで正しく数値表現

表 3.9: 定義した演算トークンに該当しない事例

(a)	日経平均、じり安　週初から値幅200 円の範囲で推移
(b)	日経平均、15 年度2448 円下落　年度ベースで 5 年ぶり

を出力できた数，できなかった数を算出する．[*6]表 3.8 に評価結果を示す．表 3.8 の算出において，「日経平均前引け、97 円安の17,137 円」のように 1 件の市況コメントに複数の数値表現が含まれている場合は，各数値表現に対して正しい数値を出力できている数を算出した．全 741 件の評価データにおいて，株価の数値表現は合計で 863 事例含まれていた．このうち，数値の演算操作の推定手法を導入した *mlp-enc* では，640 事例の数値表現を参照テキストと比較して正しく出力できていた．一方，本手法を導入していない-*num* では，92 事例の数値表現を正しく出力できていたが，その他の 771 事例において数値表現が誤っていた．以上により，数値の演算操作の推定を行うことにより，多くの事例で株価の数値表現を正しく出力できていることが確認できた．

加えて，表 3.8 における *mlp-enc* の数値表現の誤り 223 事例について分析を行った．具体的には，正解の演算操作とテキスト，および，推定した演算操作と生成テキストの四つ組を参照し，人手による分析を行った．その結果，*mlp-enc* の数値表現の誤りである 223 事例のうち 81 事例は，数値表現の切り上げ・切り捨て操作 (<OP3>，<OP5> 等) の推定誤りが起因していることが分かった．例えば，参照テキストでは演算トークンが <OP5> である数値表現について <OP3> と推定した事例が 37 事例と一番目に多く，参照テキストでは演算トークンが <OP4> である数値表現について <OP3> と推定した事例が 26 事例と 2 番目に多いことが分かった．また，その他の 142 事例については，参照テキストと生成テキストにおける内容選択の異なりに起因していることが分かった．具体例として，参照テキストでは「日経平均、反落で始まる　下げ幅は100 円超える」のように株価の下げ幅について言及しているのに対し，生成テキストでは「日経平均、反落で始まる　17600 円台」のように株価自体が何円台になったかについて言及している事例などが挙げられる．この事例の場合，参照テキストに含まれる数値 (100 円) と生成テキストの数値 (17600 円) は異なるため，表 3.8 において不正解として算出されていることに注意されたい．

また，本研究では，市況コメントで用いられる価格の差分や切り上げ・切り捨てした値などの数値表現を生成するために，12 種類の演算トークン (表 3.1) を導入している．しかし，市況コメントの生成時に正解の数値表現に対応する演算トークンが無い場合，数値表現の誤りが発生することが懸念される．そこで，本研究で提案する 12 種類の演算トークンの網羅性について調査を行った．具体的には，実験で使用した評価データにおける株価の数値表現である全 863 事例において，定義した演算トークンに該当しない数値表現の事例を人手によって確認した．その結果，定義した演算トークンに該当しない事例が 2 つあることが分かった． 表 3.9 にそれらの事例を示す． 例え

[*6] その他の評価方法として，入力の株価データの値動きと生成テキストに含まれる数値表現の比較により評価する方法が考えられる．しかし，生成テキストで言及される数値表現が入力の株価データのどの値動きに対応するかは自明ではないため自動評価が難しい．そのため本研究では人手評価において株価データの値動きと生成テキストの数値表現を比較する評価を実施する．

表 3.10: 人手評価の結果

モデル	情報性	流暢性	外部情報
Human	**95**	95	25
mlp-enc	85	93	1
baseline	28	**100**	6

ば，事例 (a) における「200 円」は，週の初めから現在までに株価がどの程度の範囲で値動きしたかを表す値幅であり，表 3.1 の演算トークンに該当するものは存在しない． また，事例 (b) における「2448 円」は，該当年度において株価がどの程度下落したかを表す年度ベースの下げ幅であり，こちらも同様に表 3.1 に該当するものは存在しない． これらの数値表現を算出するためには，対応する演算トークンを導入する必要があることに注意されたい． 以上の調査の結果，実験で使用した評価データにおいて定義した演算トークンに該当しない事例は 2 件であり， 演算トークンの網羅性が起因となる数値表現の誤り発生は少ないことが分かった．

3.4.4 人手評価結果

表 3.10 に提案手法 (*mlp-enc*)，ベースライン手法 (*baseline*) で生成した市況コメントおよび 人手で書かれた市況コメント (*Human*) に対する人手評価の結果を示す．表 3.10 において，情報性および流暢性は，0 または 1 の 2 段階評価においてスコアが 1 であった事例数，外部情報は生成テキストに外部情報を含む事例数を表している．

人手評価の結果，人手で書かれた市況コメントよりもやや劣るものの，提案手法により情報性および流暢性に関して質の高い市況コメントを生成できることが分かった．また，提案手法とベースライン手法の人手評価結果の比較では，情報性について，提案手法がベースライン手法を大幅に上回る結果となった．しかし，流暢性の観点においては，ベースライン手法が提案手法を上回っていた．

情報性に関する人手評価において提案手法がベースライン手法よりも大幅に上回っていた理由として，図 3.3 のように，提案手法はベースライン手法と比べて生成テキスト中の株価の数値表現の誤りが少ないことが主な要因として考えられる．また，ベースライン手法は数値の演算操作の推定手法を使用していないため，3.4.3 項の-*num* のように多くの事例において，数値表現の誤りが含まれていることが推察できる．これらの結果から，時系列株価データの値動きを適切に述べているかに基づいた情報性の評価において，評価者が提案手法の生成テキストの情報性が高いと判定したことが考えられる．

加えて，生成テキストの流暢性に関する人手評価においてベースライン手法が提案手法を上回っていた理由として，提案手法における演算操作の推定誤りによって，提案手法が「日経平均、上げ幅 0 円超える」といったテキストを生成していたことが原因として考えられる．「0 円超える」と

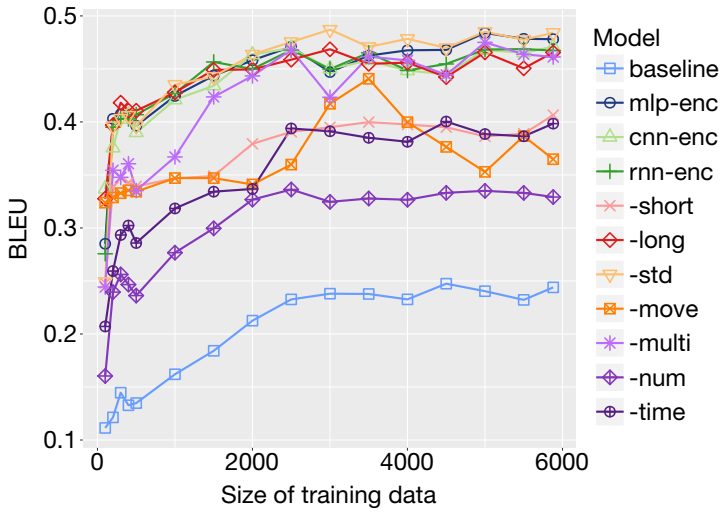


図 3.4: 各学習データサイズにおける BLEU スコアの比較

いった表現は市況コメントにおいて通常使われず不自然であり，株価の値動きについても述べられていないため，評価者がこのような表現を含む生成テキストを流暢性および情報性が低い市況コメントと判定したことが考えられる．実際に，各評価事例に対する人手評価結果を集計したところ，評価者が流暢性に 0 を付けた 7 件の提案手法の生成テキストのうち，6 件において「0 円」という表現が含まれていることが分かった．また，「0 円」という表現は通常の市況コメントでは用いられないため，価格を “単語” として予測するベースライン手法の語彙には含まれていない．そのため，ベースライン手法では上記のような「0 円超える」といった表現を含む不自然なテキストを生成しない．このことから，評価者はベースライン手法で生成された全ての市況コメントについて流暢性が高いと判定したことが考えられる．

また，人手で書かれた市況コメント (*Human*) のうち，5 件の流暢性が低いと判定された理由について確認を行った．その結果，5 件の市況コメントのいずれにおいても「日経平均、反発して始まる　42 円高、短期的な戻りを期待」といった，価格の言及方法について独特の略し方が用いられていることが分かった．その要因として，本実験では日経平均株価ニュースのヘッドラインテキストを市況コメントとして使用しており，文字数やスペース等の制約が存在することが関係していると考えられる．このような略した表現は，ほかの多くの市況コメントでは用いられないことから流暢性が低いと判定されたと考えられる．

3.4.5　各学習データサイズのモデル精度への影響

学習データサイズを変化させた場合のモデル精度への影響について分析を行った．図 3.4 に各学習データサイズにおける各モデルの BLEU スコアの比較を示す．分析では，開発データセットに対する BLEU スコアを算出した．実験結果より，学習データサイズを 3,000 事例にした時に，ほとんどのモデルにおいて BLEU スコアが飽和していることが分かった．また，各モデル間でスコ

アの収束の早さに大きな違いは無かった.

3.5 本章のまとめ

本研究では，時系列株価データから株価の値動きを概況する市況コメントを自動生成するためのData-to-Text モデルを提案した．時系列株価データを概況する市況コメントには，時系列データ中の数値への言及，過去の価格の変動との比較，テキストが書かれる時間帯によって言及する内容が異なる，などの特徴があり，本研究では大きく分けて 3 つの手法を提案した．実験では自動評価および人手評価を実施し，提案手法はベースライン手法と比べて，株価の市況コメントの特徴を捉えた正確なテキストを生成できることを示した.

第 4 章

数値気象予報からの天気予報コメントの自動生成

4.1 研究概要

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理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[*1]や Yahoo!天気[*2]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 (Belz, 2007; Goldberg et al., 1994).

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 (Liang et al., 2009; Reiter, S. Sripada, et al., 2005; S. Sripada, Reiter, and Davy, 2004)，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る．

ここで，図 4.1 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

[*1] https://weathernews.jp/
[*2] https://weather.yahoo.co.jp/weather/

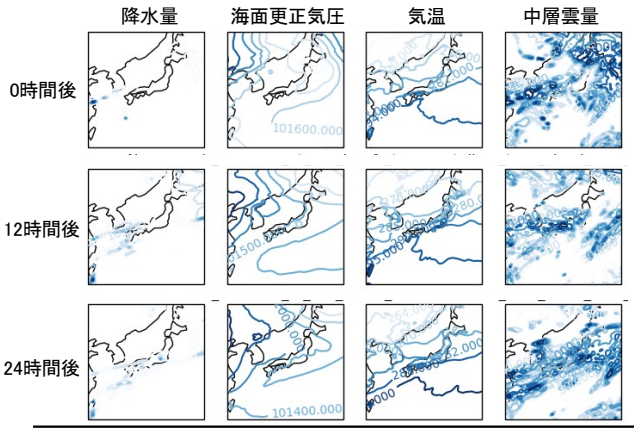
ぞれの時間変化を考慮しなければならないことである．例えば，図 4.1 では，降水量や雲量といった複数の物理量の時間変化に応じて，日差しが出た後に雲が広がり雨が降ることについて言及されている．次に，第二の問題は，天気予報コメントは，対象となる地域やコメントの配信時刻，日付といったメタ情報に基づいて記述されることである．例えば，午前中に配信される天気予報コメントでは，図 4.1 のように，配信日当日の日中から夕方にかけた天気に言及することが多く，夕方以降に配信される天気予報コメントでは，配信日当日の夜から翌日の日中の天気に言及する傾向がある．最後に，第三の問題は，天気予報サイトのユーザーは天気予報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 4.1 では，降水量，雲量，気圧など，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，MLP や 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，これらの重要な情報を選択する内容選択に向けて，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．

4.2 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．



配信日時: 4 月 6 日 午前 5:51，東京

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が段々と降り出します。外出時に雨が降っていなくても傘を持ってお出かけ下さい。

図 4.1: 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

表 4.1: 数値予報マップの概要

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

4.2.1 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 4.1 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，

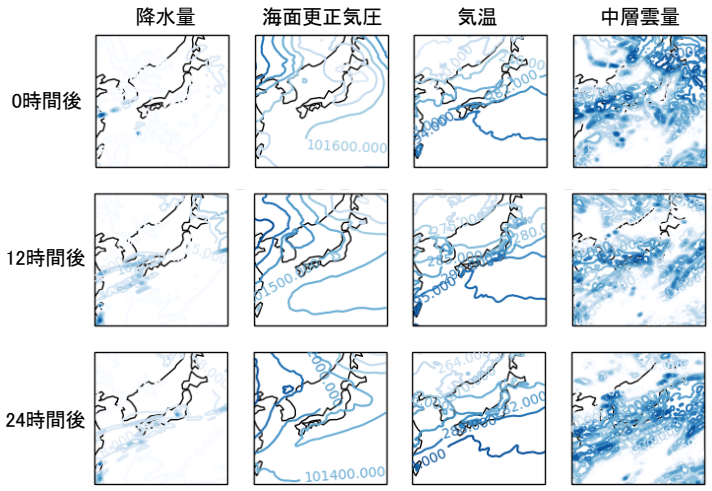
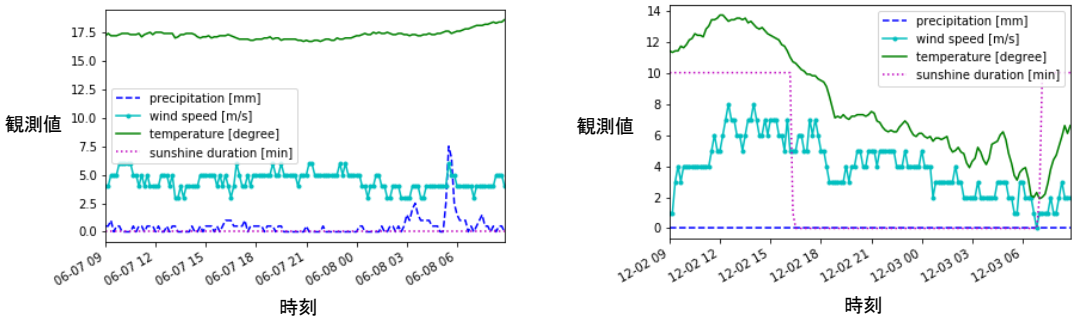


図 4.2: 数値予報マップの例

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 4.2 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 4.2 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．

4.2.2　気象観測データ



(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 4.3: 「所沢」エリアにおける 10 分毎の観測値の例

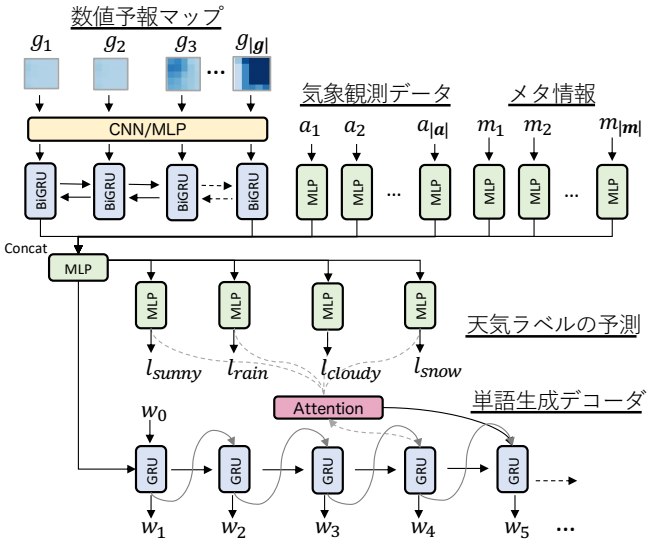


図 4.4: 提案モデルの概要

気象観測データとして，気象庁の地域気象観測システム[*3](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 4.3 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 4.3a および図 4.3b 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．

4.3 提案手法

近年，Table-to-Text タスク (Lebret et al., 2016; Mei et al., 2016b) や動画キャプション生成タスク (Long et al., 2018; Yao et al., 2015) 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 (Cho et al., 2014; Sutskever 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 4.4 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 $\boldsymbol{g} = (g_i)_{i=1}^{|\boldsymbol{g}|}$，降水量や気温などの過去の観測データ $\boldsymbol{a} = \{a_i\}_{i=1}^{|\boldsymbol{a}|}$，配信日時や対象エリア等の

[*3] https://www.jma.go.jp/jp/amedas/

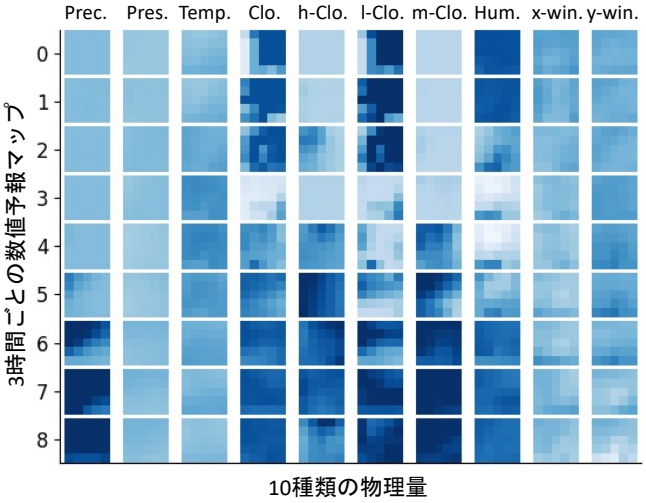


図 4.5: 日本全体の数値予報マップから抽出した東京エリア周辺の数値予報マップ

コメントに関するメタ情報 $\boldsymbol{m} = \{m_i\}_{i=1}^{|\boldsymbol{m}|}$ の 3 種類のデータを入力とし，天気予報コメント $\boldsymbol{w} = (w_i)_{i=1}^{|\boldsymbol{w}|}$ および言及すべき重要な情報を表す天気ラベル $\boldsymbol{l} = \{l_i\}_{i=1}^{|\boldsymbol{l}|}$ 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 または 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．

4.3.1 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，151×121 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 5×5 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 4.5 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 4.5 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 4.5 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．

4.3.2 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 4.5 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．

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，入力の動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ネルとして 3 チャネル (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 チャネルを用いて特徴抽出を行う．

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 5×5 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．

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:

$$h^g = [h_1^g; h_{|\boldsymbol{g}|}^g], \tag{4.1}$$

ここで，[;] はベクトルの連結演算子を表す．

4.3.3 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:

$$h^a = [h_1^a; h_2^a; \cdots; h_{|\boldsymbol{a}|}^a]. \tag{4.2}$$

4.3.4 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，午前中または午後に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 (5)，(6) のように配信時刻に依存する表現が使われる．

(5) 今日 (土)は日差しが届いても、気温は低空飛行で身体の芯まで凍える寒さ。

(6) 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 (7) のように，エリアに依存する表現が用いられる．

(7) 波も穏やかでマリンレジャーも楽しめそうです。

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 $\boldsymbol{m}$ 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:

$$h^m = [h_1^m; h_2^m; \cdots; h_{|\boldsymbol{m}|}^m]. \tag{4.3}$$

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:

$$s_0 = ReLU(MLP([h^g; h^a; h^m])). \tag{4.4}$$

4.3.5 内容選択モデルによる天気ラベルの予測

天気予報コメントにおいて，天気予報コメントの情報性とその正確さは重要である．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 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 (K. Aoki, Miyazawa, 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 (Z. 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．

表 4.2: 天気ラベルと対応する手がかり語

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

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 4.2 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[*4]．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 (8) には，手がかり語である「日差し」,「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．

(8)　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が段々と降り出します。外出時に雨が降っていなくても傘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

[*4] 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．

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．

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 ($\boldsymbol{g}$，$\boldsymbol{a}$，$\boldsymbol{m}$) 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るべきか否か判定する．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 ($\boldsymbol{g}$，$\boldsymbol{a}$，$\boldsymbol{m}$) 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．

4.3.6 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 4.4 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 (Chen et al., 2019; Wiseman et al., 2017)（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:

$$p(w_t|w_{<t}, \boldsymbol{g}, \boldsymbol{a}, \boldsymbol{m}, \boldsymbol{l}) = softmax_{w_t}(W_s s_t^w), \tag{4.5}$$

$$s_t^w = GRU(w_{t-1}, s_{t-1}^w, c_t). \tag{4.6}$$

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 $(\boldsymbol{g}, \boldsymbol{a}, \boldsymbol{m})$ と天気ラベル $\boldsymbol{l}$ 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 $\boldsymbol{l}$ 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:

$$c_t^l = \sum_{i=1}^{|\boldsymbol{l}|} \alpha_{t,i}^l s_i^l, \tag{4.7}$$

$$\alpha_{t,i}^l = \frac{exp(\eta(s_{t-1}^w, s_i^l))}{\sum_{j=1}^{|\boldsymbol{l}|} exp(\eta(s_{t-1}^w, s_j^l))}. \tag{4.8}$$

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 4.8 におけるスコア関数 η 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 4.7，4.8 と同様に導出できることに注意されたい．

4.4 実験設定

4.4.1 データセット

実験には，株式会社ウェザーニューズ[*5]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[*6]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，1326 件，1329 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 $\boldsymbol{g}$ 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

[*5] https://weathernews.jp/

[*6] http://database.rish.kyoto-u.ac.jp

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．

4.4.2 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 et al., 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 et al., 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．

4.4.3 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`[*7](Post, 2018)，ROUGE-1 の計測には `rouge`[*8]を使用した．また，ROUGE-1 のスコ

[*7] https://github.com/mjpost/sacrebleu
[*8] https://github.com/pltrdy/rouge

表 4.3: 人手評価指標

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が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．

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 4.2 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 4.3 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参照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，参照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

表 4.4: 実験に使用したモデルの概要

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	✓	–

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下さい。」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.

4.4.4 比較モデル

表 4.4 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，Z. 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，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．Z. 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，Z. Wang et al., 2020 の手法を参考に，内容一致制約損失を提案モデルへ導

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にすることで，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 4.4 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[*9]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，4.4.1 項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，上記の手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 4.4 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．

[*9] 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．

表 4.5: 各モデルの評価データに対する自動評価スコア

モデル	B-4	R-1	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

4.5 実験結果

表 4.5 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（B-4）および ROUGE スコア（R-1），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 4.5 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．

4.5.1 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 4.5 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 4.5 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．

表 4.6: 評価データに対する内容選択モデルの予測精度

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

4.5.2 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 4.5 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 4.5 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 4.6 に内容選択モデル単体の予測精度を示す．表 4.6 の内容選択モデルの精度と，表 4.5 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．

4.5.3　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 (Z. 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 4.5 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 4.5 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．

4.5.4　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 4.5 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 4.7 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 4.7 における Δ_{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 4.8 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 (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，エリアに依存表現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 4.8 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

表 4.7: 各依存表現に対する自動評価スコア

カテゴリ	依存表現	モデル (5)			モデル (6)			Δ_{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

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．

4.5.5 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 4.5 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 4.5 の BLEU および

表 4.8: 各依存表現の各データにおける出現回数

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	9,910	5,176	5,203
	今日	19,328	9,636	9,521
	午前	1,301	559	494
	午後	4,628	2,280	2,314
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	1,732	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

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

表 4.9: 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

モデル	B-4	R-1	晴れ			雨			曇り			雪		
			P%	R%	F_1%	P%	R%	F_1%	P%	R%	F_1%	P%	R%	F_1%
$(1)_{max}$	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 4.10: 事例ベース推論に基づく手法の生成テキスト

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じる一日に。
$(1)_{max}$	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。

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 4.5 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 4.9 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 4.10 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 4.10 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

表 4.11: 人手評価の結果

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

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，参照テキストと同様に，雨と曇り空について言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，参照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参照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．

4.5.6 人手評価結果

表 4.11 に人手評価の結果を示す．表 4.11 における，† 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 4.1 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，統計的有意に情報性が優れていることが分かった．これは，表 4.5 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 (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．この原因として 2 つの可能性が考えられる．まず 1 つ目は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という可能性である．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考えられる (Iso et al., 2019; Wiseman et al., 2017)．次に 2 つ目は，内容選択モデ

表 4.12: 「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 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。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。
(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。

ルに対する注意機構が単語生成デコーダに悪影響を与えたという可能性である．提案モデル (6) では，重要な情報を明示的に記述することを期待して天気ラベルに対する注意機構を導入したものの，単語生成デコーダで生成される単語の中には気象情報とは関係性の低い助詞や接続詞等の単語も含まれる．すなわち，テキストの一貫性や文法性に関わるこれらの単語の生成時に注意機構を介して気象情報に関する追加情報が与えられたことで，結果的に生成テキストの一貫性が劣化したことが推察できる．この問題の対策として，単語生成デコーダに予測した天気ラベルを導入する方法の改善が考えられる．例えば，気象情報に関する内容語（晴れ, 曇り等）の生成時に限り天気ラベルに対する注意機構を有効にする方法などが考えられる．

また，モデル (6) の生成テキストでは，以下に示す事例 (9) 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[*10]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．

(9) 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 4.12，4.13，4.14 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 4.12，4.13，4.14 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 4.12 では，各モデルの生成テキストにおいて参照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参照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 4.13 では，各モデルの生成テキストでは日差しや

[*10] 一貫性の平均スコアは 1.8

表 4.13: 「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

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ます。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下さい。

表 4.14: 「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

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。

気温の低下について言及できている．一方，モデル (6) は参照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 4.13 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 4.14 では，モデル (6) は参照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，参照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，参照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，以下の参照テキスト (10) と生成テキスト (11) 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，参照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．

(10) 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。

(11) 今日 (木) は日差しが届いても、ニワカ雨の可能性があります。お出かけの際は折りたたみ傘があると安心です。

4.6 本章のまとめ

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，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，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，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を実施し，提案手法はベースライン手法と比べて，天気予報コメントの特徴を捉えた正確なテキストを生成できることを示した．

第 5 章

結論と今後の課題

5.1 結論

本研究では，Data-to-Text モデルにおける生成テキストの正確性の問題に対して，Data-to-Text モデルの高度化を通して貢献した．

1 つ目の研究では，時系列株価データから市況コメントを自動生成するための Data-to-Text モデルを提案した．時系列株価データを概況する市況コメントには，過去の価格の変動との比較，テキストが書かれる時間帯によって言及する内容が異なるなどの特徴がある．また，市況コメントでは時系列データ中の株価に言及することもあれば，株価の増減幅のように時系列データから演算操作によって算出された数値に言及することもある．本研究ではこれらの特徴に対して，まず，株価の短期的・長期的な変化を捉えるためのエンコード手法を検証し，テキストが書かれる時間帯に依存する表現を生成するために時間帯情報の導入に取り組んだ．また，市況コメントで言及される株価の数値表現を生成する方法として，数値の演算操作を推定し計算することで数値表現を生成する手法を提案した．これにより，株価の増減幅といった演算操作が必要な数値表現についても正確に言及することが可能となった．実験では自動評価および人手評価を実施した．自動評価では，生成テキストと参照テキストの単語一致に基づく BLEU，および，「続落，前引け」などの株価の変動や時間帯に依存する表現を正しく出力でき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大幅に性能向上することを確認した．さらに人手評価では，流暢性と情報性の観点において，提案手法により株価の市況コメントの特徴を捉えたテキストを生成できることを示した．以上の結果より，時系列株価データから市況コメントを正確にテキスト化するための提案手法の有用性を確認することができた．

2 つ目の研究では，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，複数の物理量からなる時系列数値データを考慮する必要がある，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といった特徴がある．また，天気予報コメントにおいて「晴れ」「雨」「曇り」「雪」といった気象情報は重要であり，適切かつ明示的に言及することが求められる．本研究ではこれらの特徴に対して，まず，数値予報マップのエンコード手法を検証し，配信時刻やエリア名等の表現を

生成するためにメタ情報の導入に取り組んだ．さらに，入力データ中の重要な情報を予測するための内容選択モデルを導入し，これらを生成時に考慮することで気象情報に適切に言及する方法を提案した．本研究では，手がかり語に基づいて出力テキスト側から気象情報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した．これにより，単語の表層に基づいて入出力データの対応が取れない場合であっても，入力データから気象情報を予測する内容選択モデルを導入することが可能となった．実験では，自動評価および人手評価を実施した．自動評価ではまず，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，天気ラベルに関する適合率，再現率，F_1 スコアを用い，提案手法がベースライン手法よりも性能向上することを確認した．また，配信時刻やエリア名等のメタ情報を導入することで，メタ情報に依存する表現をより適切に言及できることを示した．さらに人手評価では，入力データに対して内容選択を行う提案手法は，内容選択を行わないベースライン手法に比べて，生成テキストの情報性の観点で優れていることを確認した．以上の結果より，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正確にテキスト化するための提案手法の有用性を確認することができた．

5.2 今後の課題

本節では，まず本研究で取り組んだ 2 つの研究について今後の課題を述べる．その後，2 つの研究を通して気づきを得られた Data-to-Text 生成課題全体における今後の方向性を議論する．

5.2.1 時系列株価データからの市況コメントの自動生成

1 つ目の研究として取り組んだ株価の市況コメント生成について，今後の課題を 4 点挙げる．

1 点目の課題として，3.4.3 項で議論した本研究で提案した演算トークンの網羅性の課題が挙げられる．前述の通り，提案手法では市況コメントの生成時に適切な演算操作が定義されていない場合，数値表現の誤りが発生する可能性がある．この課題は生成テキストの正確性の低下に繋がることから，生成システムの実用化の観点においては，事前に定義されていない演算操作であっても適切に数値に言及できることが望ましい．そのため，本課題に対する対策として，入力データや生成テキスト等の文脈に基づいてモデルが適切な演算操作を導出し，その結果に基づいて数値表現に言及する手法などが必要であると考えられる．

次に 2 点目の課題として，いつからいつまでの株価の値動きや上げ幅について言及するべきかモデルに考慮させることが考えられる．例えば，実際の日経平均株価の上げ幅が 300 円だったにも拘わらず，現在のモデルでは「日経平均、上げ幅 200 円超える」といった市況コメントを生成することがある．このような生成テキストは，誤りではないが正確な市況コメントとは言えない．この問題を解決するために，モデルが市況コメントを生成する際に，生成テキストで言及する時系列株価データの期間を選択するための機構が必要であると考えられる．

3 点目の課題として，時系列株価データのエンコード手法のさらなる検討のために，近年さまざまな言語処理タスクで有用性が示されている Self-Attention Network (SAN) をエンコーダとして

利用することが挙げられる．SAN の利点の 1 つとして，RNN などの再帰構造のネットワークと比べて，時系列データにおける各タイムステップ間の依存関係をより短い距離で考慮できることが挙げられる (Vaswani et al., 2017)．これにより，全 62 タイムステップからなる 1 日分の時系列株価データといった長い系列長の時系列データにおける長距離の依存関係をモデル化する能力が向上し，短期的及び長期的な株価の変化を捉える性能が改善することが期待できる．

4 点目の課題として，株式市場全体の個別銘柄などの複数の時系列データを考慮したより高度な市況コメント生成に取り組むことが考えられる．本研究では，日経平均株価を例に単一の時系列データに対する市況コメントを扱っていたため，一般的に金融アナリストから報告されるマーケット動向[*1]等と比べると，時系列データの変動やデータ自体は複雑ではなく，市況コメント自体も比較的単調な言い回しや数値表現に限られている傾向があった．一方，Data-to-Text モデルの実用化の観点においては，本研究で扱ったような市況コメントだけではなく，金融アナリストの市場分析等の業務サポートにも繋がるより高度な市況コメント生成の需要が高いことが考えられる．したがって今後の課題として，演算操作や数値データのエンコード手法等の本研究で得られた知見を踏まえ，複数の時系列データ等を対象としたより高度な市況コメント生成に取り組むことが考えられる．

5.2.2　数値気象予報からの天気予報コメントの自動生成

次に 2 つ目の研究である天気予報コメント生成について，今後の課題を 4 点挙げる．

まず 1 点目の課題として，本研究で取り組んだ，内容選択モデルの導入により生成テキストの正確性を改善する試みには依然として改良の余地が残っている点を挙げる．すなわち，4.5.2 項で議論した，現状のモデルでは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できていないという課題である．この課題は，内容選択モデルが重要な情報を正確に選択できていたとしても，生成テキストにおいて誤った情報の混入や情報の不足等が生じる懸念があるため解決が求められる．この課題の対策として，例えば，テキスト生成時に内容選択モデルの予測結果を必ず反映するための制約の導入などが考えられる．

2 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．

3 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が，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

[*1] https://www.nomura.co.jp/market/conditions/

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．

4 点目の課題として，入力の気象データとして日本周辺全体の数値予報マップを用いることが挙げられる．本研究では，各エリアに対する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，日本周辺全体の数値予報マップから抽出した各エリアの数値予報マップのみを使用している．しかし，実際の天気予報コメントの制作過程では，対象エリア周辺といった局所的な気象情報だけでなく，日本周辺全体の大域的な気象情報を参照している．したがって，生成モデルにおいても同様に大域的な気象情報を考慮することができれば，より情報性の高い天気予報コメントを生成できることが考えられる．例えば，台風や長時間継続するような雨に関する予報は，その対象エリア周辺の気象情報だけでなく大域的な雲や前線の動きを考慮することで，より適切に言及できることが期待できる．

5.2.3 Data-to-Text 生成課題における今後の方向性

本研究では，市況コメントおよび天気予報コメントの自動生成タスクに焦点をあて，数値データの特徴を捉えるためのモデルの高度化を通して品質の高いテキスト生成が実現できることを示した．一方で，Data-to-Text 生成技術の最終的な目標である実用化に向けては解決すべき課題がいくつか残っている．そこで本研究で取り組んだ 2 つの研究を通して気づきを得られた Data-to-Text 生成課題における今後の研究の方向性として「外部情報の考慮」を例に挙げ，議論する．

これまで Data-to-Text 生成課題では様々なドメインを対象とした研究が取り組まれているが，実世界におけるテキストの多くは主眼を置く情報に含まれる事実だけでなく，過去の記録や世界知識等の外部情報も考慮した上で記述されている．例えば日経平均株価の市況コメントでは，株価の値動きに関する情報に加えて「米株安や円高で、下げ幅 100 円超える」のように外国株や円高の情報や，「日経平均、15 年度 2448 円下落　年度ベースで 5 年ぶり」（表 3.9）のように過去の記録に基づく情報が言及される．また天気予報コメントでは，「クリスマス、年末の準備もはかどる天気です。」などのイベント情報や，「明日は台風 7 号が接近」のように過去の台風の発生数を踏まえた内容が言及される．これらのドメイン以外でも同様であり，例えば野球の試合要約では試合内容に関する情報に加えて「ロッテの佐々木朗希が28 年ぶり史上 16 人目、完全試合を達成」のように過去の記録に基づく情報が頻繁に言及されている．

これらの特徴を踏まえると Data-to-Text 生成技術の実用化に向けて，本研究で取り組んだ入力データの事実について正確に言及するいう方向性の他に，前述のような過去の記録や世界知識等の外部情報を考慮した生成手法の確立という研究の方向性が考えられる．具体的には，外部情報に言及する際に参照するデータベースの整備，文脈に応じてデータベース上から適切な知識を獲得し言及する手法，生成された外部情報自体の正確性の評価などが今後の方向性として考えられる．

謝辞

本学位論文は著者が東京工業大学工学院情報通信系情報通信コースに在籍中の研究成果をまとめたものです．本研究を遂行するにあたり，多くの方のご支援，ご指導をいただきました．

指導教員である奥村学教授には，修士課程から計 5 年間に渡り，本研究の構想から論文執筆に至るまで終始熱心かつ丁寧なご指導とご鞭撻を賜りました．研究を進めていく中で，研究者としての心構え，研究の楽しさなど多くの事を学ばせていただき，博士課程への進学や研究者としてキャリアを始めるきっかけとなりました．ここに深謝の意を表します．

船越孝太郎准教授，奈良先端科学技術大学院大学の上垣外英剛准教授，産業技術総合研究所の高村大也研究チーム長，名古屋大学の笹野遼平准教授には，研究に対する姿勢や研究の方向性，論文執筆等について様々なご指導，ご助言を頂きました．深く感謝いたします．

本論文の審査を引き受けてくださった熊澤逸夫教授，中山実教授，篠崎隆宏准教授，北陸先端科学技術大学院大学の白井清昭准教授に感謝申し上げます．先生方には多角的な視点から貴重なご助言，ご指導をいただくことができました．

研究室の皆様には大変お世話になりました．特に研究室秘書の飯山信子氏にはあらゆる事務処理を引き受けてくださり，研究室運営において間違いなくかけがえの無い存在でした．あらためて感謝申し上げます．

産業技術総合研究所のテクニカルスタッフとして勤務していた際にお世話になった渡邉亮彦氏，宮澤彬氏，五島圭一氏，柳瀬利彦氏，宮尾祐介教授に感謝申し上げます．研究室を離れ，初めて共同で取り組んだ産総研での研究は，私にとって充実したかけがえのない経験になっています．

本研究を進めるにあたり株式会社ウェザーニューズ様からデータ提供のご協力をいただきました．同社の萩行正嗣氏には，データの使用方法や研究に関するご助言をいただきました．心から感謝致します．

最後に，これまで私を温かく応援してくれた両親と姉妹，私をいつも明るく励まし続けてくれた妻に心から感謝します．

参考文献

Adeyanju, Ibrahim (2012). “Generating Weather Forecast Texts with Case Based Reasoning”. In: *International Journal of Computer Applications* 45, pp. 35–40.

Angeli, Gabor, Percy Liang, and Dan Klein (2010). “A Simple Domain-Independent Probabilistic Approach to Generation”. In: *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pp. 502–512.

Aoki, Kasumi and Ichiro Kobayashi (2016). “Linguistic Summarization Using A Weighted N-gram Language Model Based on the Similarity of Time-series Data”. In: *Proceedings of IEEE International Conference on Fuzzy Systems*, pp. 595–601.

Aoki, Kasumi, Akira Miyazawa, Tatsuya Ishigaki, Tatsuya Aoki, Hiroshi Noji, Keiichi Goshima, Ichiro Kobayashi, Hiroya Takamura, and Yusuke Miyao (2019). “Controlling Contents in Data-to-Document Generation with Human-Designed Topic Labels”. In: *Proceedings of the 12th International Conference on Natural Language Generation*. Association for Computational Linguistics, pp. 323–332.

Aoki, Tatsuya, Akira Miyazawa, Tatsuya Ishigaki, Keiichi Goshima, Kasumi Aoki, Ichiro Kobayashi, Hiroya Takamura, and Yusuke Miyao (2018). “Generating Market Comments Referring to External Resources”. In: *Proceedings of the 11th International Conference on Natural Language Generation*. Association for Computational Linguistics, pp. 135–139.

Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio (2015). “Neural Machine Translation by Jointly Learning to Align and Translate”. In: *Proceedings of the 3rd International Conference on Learning Representations*.

Banaee, Hadi, Mobyen Uddin Ahmed, and Amy Loutfi (2013a). “A Framework for Automatic Text Generation of Trends in Physiological Time Series Data”. In: *Processing of IEEE International Conference on Systems, Man, and Cybernetics*, pp. 3876–3881.

Banaee, Hadi, Mobyen Uddin Ahmed, and Amy Loutfi (2013b). “Towards NLG for Physiological Data Monitoring with Body Area Networks”. In: *Proceedings of the 14th European Workshop on Natural Language Generation*. Association for Computational Linguistics, pp. 193–197.

Barzilay, Regina and Mirella Lapata (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*. Association for Computational Linguistics, pp. 331–338.

Belz, Anja (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*. Association for Computational Linguistics, pp. 164–171.

Chen, Shuang, Jinpeng Wang, Xiaocheng Feng, Feng Jiang, Bing Qin, and Chin-Yew Lin (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*. Association for Computational Linguistics, pp. 3022–3032.

Cho, Kyunghyun, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio (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*. Association for Computational Linguistics, pp. 1724–1734.

Chung, Junyoung, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio (2014). "Empirical evaluation of gated recurrent neural networks on sequence modeling". In: *NIPS 2014 Workshop on Deep Learning, December 2014*.

Cun, Y. Le, B. Boser, J. S. Denker, R. E. Howard, W. Habbard, L. D. Jackel, and D. Henderson (1990). "Advances in Neural Information Processing Systems 2". In: ed. by David S. Touretzky. Morgan Kaufmann Publishers Inc. Chap. Handwritten Digit Recognition with a Back-propagation Network, pp. 396–404. ISBN: 1-55860-100-7.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (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*. Association for Computational Linguistics, pp. 4171–4186.

Dhingra, Bhuwan, Manaal Faruqui, Ankur Parikh, Ming-Wei Chang, Dipanjan Das, and William Cohen (2019). "Handling Divergent Reference Texts when Evaluating Table-to-Text Generation". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, pp. 4884–4895.

Duboue, Pablo A. and Kathleen R. McKeown (2001). "Empirically Estimating Order Constraints for Content Planning in Generation". In: *Proceedings of the 39th Annual Meeting of*

the Association for Computational Linguistics. Association for Computational Linguistics, pp. 172–179.

Duboue, Pablo Ariel and Kathleen R. McKeown (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.

Dušek, Ondřej, David M. Howcroft, and Verena Rieser (2019). "Semantic Noise Matters for Neural Natural Language Generation". In: *Proceedings of the 12th International Conference on Natural Language Generation*. Association for Computational Linguistics, pp. 421–426.

Filippova, Katja (2020). "Controlled Hallucinations: Learning to Generate Faithfully from Noisy Data". In: *Findings of the Association for Computational Linguistics: EMNLP 2020*. Association for Computational Linguistics, pp. 864–870.

Freitas, Fabio D., Alberto F. De Souza, and Ailson R. de Almeida (2009). "Prediction-Based Portfolio Optimization Model Using Neural Networks". In: *Neurocomputing* 72.10, pp. 2155–2170.

Gatt, Albert and Emiel Krahmer (2018). "Survey of the State of the Art in Natural Language Generation: Core Tasks, Applications and Evaluation". In: *Journal of Artificial Intelligence Research* 61.1, pp. 65–170. ISSN: 1076-9757.

Gkatzia, Dimitra, Helen Hastie, and Oliver Lemon (2014). "Comparing Multi-label Classification with Reinforcement Learning for Summarisation of Time-series Data". In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, pp. 1231–1240.

Glorot, Xavier, Antoine Bordes, and Yoshua Bengio (2011). "Deep Sparse Rectifier Neural Networks". In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. Ed. by Geoffrey Gordon, David Dunson, and Miroslav Dudík. Vol. 15. Proceedings of Machine Learning Research. PMLR, pp. 315–323.

Goldberg, Eli, Norbert Driedger, and Richard I. Kittredge (1994). "Using Natural-Language Processing to Produce Weather Forecasts". In: *IEEE Expert* 9.2, pp. 45–53.

Gulcehre, Caglar, Sungjin Ahn, Ramesh Nallapati, Bowen Zhou, and Yoshua Bengio (2016). "Pointing the Unknown Words". In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, pp. 140–149.

Hochreiter, Sepp and Jürgen Schmidhuber (1997). "Long Short-Term Memory". In: *Neural Comput.* 9.8, pp. 1735–1780. ISSN: 0899-7667.

Iso, Hayate, Yui Uehara, Tatsuya Ishigaki, Hiroshi Noji, Eiji Aramaki, Ichiro Kobayashi, Yusuke Miyao, Naoaki Okazaki, and Hiroya Takamura (2019). "Learning to Select, Track, and Generate for Data-to-Text". In: *Proceedings of the 57th Annual Meeting of the Associ-*

ation for Computational Linguistics. Association for Computational Linguistics, pp. 2102–2113.

Kale, Mihir and Abhinav Rastogi (2020). "Text-to-Text Pre-Training for Data-to-Text Tasks". In: *Proceedings of the 13th International Conference on Natural Language Generation*. Association for Computational Linguistics, pp. 97–102.

Kerpedjiev, Stephan M. (1992). "Automatic Generation of Multimodal Weather Reports from Datasets". In: *Third Conference on Applied Natural Language Processing*. Association for Computational Linguistics, pp. 48–55.

Kingma, Diederik P. and Jimmy Ba (2015). "Adam: A Method for Stochastic Optimization". In: *Proceedings of the 3rd International Conference on Learning Representations*.

Kittredge, R., A. Polguere, and E. Goldberg (1986). "Synthesizing Weather Forecasts From Formatted Data". In: *Proceedings of the 11th International Conference on Computational Linguistics*.

Koehn, Philipp (2004). "Statistical Significance Tests for Machine Translation Evaluation". In: *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pp. 388–395.

Kukich, Karen (1983). "Design of A Knowledge-Based Report Generator". In: *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, pp. 145–150.

Lebret, Rémi, David Grangier, and Michael Auli (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*. Association for Computational Linguistics, pp. 1203–1213.

Li, Jiwei, Michel Galley, Chris Brockett, Georgios Spithourakis, Jianfeng Gao, and Bill Dolan (2016). "A Persona-Based Neural Conversation Model". In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, pp. 994–1003.

Li, Wei, Wenhao Wu, Moye Chen, Jiachen Liu, Xinyan Xiao, and Hua Wu (2022). "Faithfulness in Natural Language Generation: A Systematic Survey of Analysis, Evaluation and Optimization Methods". In: arXiv.

Liang, Percy, Michael Jordan, and Dan Klein (2009). "Learning Semantic Correspondences with Less Supervision". In: *Proceedings of Association for Computational Linguistics and International Joint Conference on Natural Language Processing*. Association for Computational Linguistics, pp. 91–99.

Lin, Chin-Yew (2004). "ROUGE: A Package for Automatic Evaluation of Summaries". In: *Text Summarization Branches Out*. Association for Computational Linguistics, pp. 74–81.

Liu, Tianyu, Kexiang Wang, Lei Sha, Baobao Chang, and Zhifang Sui (2018). "Table-to-Text Generation by Structure-Aware Seq2seq Learning". In: *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, pp. 4881–4888.

Long, Xiang, Chuang Gan, and Gerard de Melo (2018). "Video Captioning with Multi-Faceted Attention". In: *Transactions of the Association for Computational Linguistics* 6, pp. 173–184.

Luong, Thang, Ilya Sutskever, Quoc Le, Oriol Vinyals, and Wojciech Zaremba (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*. Association for Computational Linguistics, pp. 11–19.

Ma, Shuming, Pengcheng Yang, Tianyu Liu, Peng Li, Jie Zhou, and Xu Sun (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*. Association for Computational Linguistics, pp. 2047–2057.

Mei, Hongyuan, Mohit Bansal, and Matthew R. Walter (2016a). "Listen, Attend, and Walk: Neural Mapping of Navigational Instructions to Action Sequences". In: *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pp. 2772–2778.

Mei, Hongyuan, Mohit Bansal, and Matthew R. Walter (2016b). "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*. Association for Computational Linguistics, pp. 720–730.

Mikolov, Tomáš, Martin Karafiát, Lukáš Burget, Jan Černocký, and Sanjeev Khudanpur (2010). "Recurrent Neural Network Based Language Model". In: *Proceedings of the 11th Annual Conference of the International Speech Communication Association*. International Speech Communication Association, pp. 1045–1048.

Narayan, Shashi, Shay B. Cohen, and Mirella Lapata (2018). "Don't Give Me the Details, Just the Summary! Topic-Aware Convolutional Neural Networks for Extreme Summarization". In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pp. 1797–1807.

Nie, Feng, Jin-Ge Yao, Jinpeng Wang, Rong Pan, and Chin-Yew Lin (2019). "A Simple Recipe towards Reducing Hallucination in Neural Surface Realisation". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, pp. 2673–2679.

Novikova, Jekaterina, Ondrej Dušek, and Verena Rieser (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*.

Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu (2002). "BLEU: A Method for Automatic Evaluation of Machine Translation". In: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, pp. 311–318.

Parikh, Ankur, Xuezhi Wang, Sebastian Gehrmann, Manaal Faruqui, Bhuwan Dhingra, Diyi Yang, and Dipanjan Das (2020). "ToTTo: A Controlled Table-To-Text Generation Dataset". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, pp. 1173–1186.

Portet, François, Ehud Reiter, Albert Gatt, Jim Hunter, Somayajulu Sripada, Yvonne Freer, and Cindy Sykes (2009). "Automatic Generation of Textual Summaries from Neonatal Intensive Care Data". In: *Artificial Intelligence* 173.7-8, pp. 789–816.

Post, Matt (2018). "A Call for Clarity in Reporting BLEU Scores". In: *Proceedings of the Third Conference on Machine Translation: Research Papers*. Association for Computational Linguistics, pp. 186–191.

Puduppully, Ratish, Li Dong, and Mirella Lapata (2019). "Data-to-Text Generation with Content Selection and Planning". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 33, pp. 6908–6915.

Radford, Alec, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever (2019). "Language Models are Unsupervised Multitask Learners". In.

Raffel, Colin, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu (2020). "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer". In: *Journal of Machine Learning Research* 21.140, pp. 1–67.

Reiter, Ehud and Robert Dale (1997). "Building applied natural language generation systems". In: *Natural Language Engineering* 3.1, pp. 57–87.

Reiter, Ehud, Somayajulu Sripada, Jim Hunter, Jin Yu, and Ian Davy (2005). "Choosing Words in Computer-Generated Weather Forecasts". In: *Artificial Intelligence* 167.1-2, pp. 137–169.

Rush, Alexander M., Sumit Chopra, and Jason Weston (2015). "A Neural Attention Model for Abstractive Sentence Summarization". In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pp. 379–389.

See, Abigail, Peter J. Liu, and Christopher D. Manning (2017). "Get To The Point: Summarization with Pointer-Generator Networks". In: *Proceedings of the 55th Annual Meeting of*

the Association for Computational Linguistics. Association for Computational Linguistics, pp. 1073–1083.

Sha, Lei, Lili Mou, Tianyu Liu, Pascal Poupart, Sujian Li, Baobao Chang, and Zhifang Sui (2018a). "Order-Planning Neural Text Generation From Structured Data". In: *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, pp. 5414–5421.

Sha, Lei, Lili Mou, Tianyu Liu, Pascal Poupart, Sujian Li, Baobao Chang, and Zhifang Sui (2018b). "Order-Planning Neural Text Generation from Structured Data". In: *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*. AAAI'18/IAAI'18/EAAI'18. AAAI Press.

Sharma, Mandar, Ajay Gogineni, and Naren Ramakrishnan (2022). *Innovations in Neural Data-to-text Generation*. arXiv: 2207.12571 [cs.CL].

Song, Kaitao, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu (2019). "MASS: Masked Sequence to Sequence Pre-training for Language Generation". In: *International Conference on Machine Learning*, pp. 5926–5936.

Sripada, Somayajulu, Ehud Reiter, and Ian Davy (2004). "SumTime-Mousam: Configurable marine weather forecast generator". In: *Expert Update* 6.

Sripada, Somayajulu, Ehud Reiter, Jim Hunter, and Jin Yu (2003). "Exploiting a parallel TEXT - DATA corpus". In: *Proceedings of Corpus Linguistics*.

Sripada, Somayajulu G., Ehud Reiter, Jim Hunter, Jin Yu, and Ian P. Davy (2002). "Modelling the Task of Summarising Time Series Data Using KA Techniques". In: *Applications and Innovations in Intelligent Systems IX*. Springer London, pp. 183–196.

Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le (2014). "Sequence to Sequence Learning with Neural Networks". In: *Proceedings of the 27th International Conference on Neural Information Processing Systems*, pp. 3104–3112.

Tian, Ran, Shashi Narayan, Thibault Sellam, and Ankur P Parikh (2019). "Sticking to the Facts: Confident Decoding for Faithful Data-to-Text Generation". In.

Tu, Zhaopeng, Zhengdong Lu, Yang Liu, Xiaohua Liu, and Hang Li (2016). "Modeling Coverage for Neural Machine Translation". In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, pp. 76–85.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin (2017). "Attention is All you Need". In: *Advances in Neural Information Processing Systems 30*. Ed. by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett. Curran Associates, Inc., pp. 5998–6008.

Vijayarani, S, Ms J Ilamathi, and Ms Nithya (2015). "Preprocessing techniques for text mining-an overview". In: *International Journal of Computer Science & Communication Networks* 5.1, pp. 7–16.

Vinyals, Oriol, Alexander Toshev, Samy Bengio, and Dumitru Erhan (2015). "Show and Tell: A Neural Image Caption Generator". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3156–3164.

Wang, Chaojun and Rico Sennrich (2020). "On Exposure Bias, Hallucination and Domain Shift in Neural Machine Translation". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, pp. 3544–3552.

Wang, Zhenyi, Xiaoyang Wang, Bang An, Dong Yu, and Changyou Chen (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*. Association for Computational Linguistics, pp. 1072–1086.

Wilcoxon, Frank (1945). "Individual Comparisons by Ranking Methods". In: *Biometrics Bulletin* 1.6, pp. 80–83.

Wiseman, Sam, Stuart Shieber, and Alexander Rush (2017). "Challenges in Data-to-Document Generation". In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pp. 2253–2263.

Yao, Li, Atousa Torabi, Kyunghyun Cho, Nicolas Ballas, Christopher Pal, Hugo Larochelle, and Aaron Courville (2015). "Describing Videos by Exploiting Temporal Structure". In: *Proceedings of the 2015 IEEE International Conference on Computer Vision*. IEEE Computer Society, pp. 4507–4515.

Zhang, Dell, Jiahao Yuan, Xiaoling Wang, and Adam Foster (2018). "Probabilistic Verb Selection for Data-to-Text Generation". In: *Transactions of the Association for Computational Linguistics* 6, pp. 511–527.

Zhang, G. Peter and Min Qi (2005). "Neural Network Forecasting for Seasonal and Trend Time Series". In: *European journal of operational research* 160.2, pp. 501–514.

研究業績

本論文に関連する業績

論文誌

- 村上聡一朗，田中天，萩行正嗣，上垣外英剛，船越孝太郎，高村大也，奥村学，“数値気象予報からの天気予報コメントの自動生成”，自然言語処理，28-4，pp.1210 - 1246
- 村上聡一朗，渡邉亮彦，宮澤彬，五島圭一，柳瀬利彦，高村大也，宮尾祐介，“時系列株価データからの市況コメントの自動生成”，自然言語処理，27-2，pp.299 - 328

国際会議論文

- Soichiro Murakami，Sora Tanaka，Masatsugu Hangyo，Hidetaka Kamigaito，Kotaro Funakoshi，Hiroya Takamura and Manabu Okumura，“Generating Weather Comments from Meteorological Simulations”，In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL2021)
- Soichiro Murakami，Akihiko Watanabe，Akira Miyazawa，Keiichi Goshima，Toshihiko Yanase, Hiroya Takamura and Yusuke Miyao，“Learning to Generate Market Comments from Stock Prices”，In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL2017)

国内会議論文

- 村上聡一朗，笹野遼平，高村大也，奥村学，“数値予報マップからの天気予報コメントの自動生成”，言語処理学会 第 23 回年次大会
- 村上聡一朗，渡邉亮彦，宮澤彬，五島圭一，柳瀬利彦，高村大也，宮尾祐介，“時系列数値データからの概況テキストの自動生成”，言語処理学会 第 23 回年次大会

その他の業績

国際会議論文

- Soichiro Murakami, Peinan Zhang, Sho Hoshino, Hidetaka Kamigaito, Hiroya Takamura and Manabu Okumura, “Aspect-based Analysis of Advertising Appeals for Search Engine Advertising”, In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies: Industry Track Papers (NAACL-HLT 2022 Industry Track Papers)
- Soichiro Murakami, Makoto Morishita, Tsutomu Hirao, Masaaki Nagata, “NTT’s Machine Translation Systems for WMT19 Robustness Task”, First Place in the En-Ja Track, The 4th Conference on Machine Translation (WMT19)

国内会議論文

- 村上聡一朗，星野翔，張培楠，“広告文自動生成に関する最近の研究動向”，2022 年度 人工知能学会全国大会 (第 36 回)
- 村上聡一朗，星野翔，張培楠，上垣外英剛，高村大也，奥村学，“LP-to-Text: マルチモーダル広告文生成”，言語処理学会 第 28 回年次大会
- 村上聡一朗，田中天，萩行正嗣，上垣外英剛，高村大也，奥村学，“Data-to-Text モデルにおけるトピック系列を用いた一貫性の制御”，2020 年度 人工知能学会全国大会 (第 34 回)
- 村上聡一朗，松岡保静，内田渉，礒田佳徳，森下睦，平尾努，永田昌明，“自然発話に頑健な機械翻訳の検討”，言語処理学会 第 25 回年次大会 (2019.3)
- 柳瀬利彦，柳井孝介，丹羽芳樹，村上聡一朗，渡邉亮彦，宮澤彬，五島圭一，高村大也，宮尾祐介，中田亨，“企業経営における意思決定支援のためのイベント抽出”，2017 年度 人工知能学会全国大会 (第 31 回)
- 渡邉亮彦，村上聡一朗，宮澤彬，五島圭一，柳瀬利彦，高村大也，宮尾祐介，“TRF: テキストの読みやすさ解析ツール”，言語処理学会 第 23 回年次大会 (2017.3)
- 村上聡一朗，笹野遼平，高村大也，奥村学，“打者成績からのイニング速報の自動生成”，言語処理学会 第 22 回年次大会

受賞

- **言語処理学会 2021 年度論文賞**，村上聡一朗，田中天，萩行正嗣，上垣外英剛，船越孝太郎，高村大也，奥村学，“数値気象予報からの天気予報コメントの自動生成”，自然言語処理，28-4，pp.1210 - 1246
- **若手奨励賞**，村上聡一朗，渡邉亮彦，宮澤彬，五島圭一，柳瀬利彦，高村大也，宮尾祐介，“時系列数値データからの概況テキストの自動生成”，言語処理学会 第 23 回年次大会