

# タスク指向対話システムにおけるユーザーの特徴を 考慮した話題導入

松原 真弓<sup>1,a)</sup> 麻場 直喜<sup>1</sup> 内藤 昭一<sup>1</sup> 川村 晋太郎<sup>1</sup> 井口 慎也<sup>1</sup> 能勢 将樹<sup>1</sup> 岡崎 直観<sup>2</sup>

受付日 2022年1月18日, 採録日 2023年4月21日

**概要:** 深層学習をはじめとする AI 技術の発展により, タスク指向対話システムを業務に活用するシーンが増えている. システムとユーザーがより協調して業務を推進するには, ユーザーが過去に話した内容を踏まえてタスクを遂行する必要がある. そこで今回, 過去にユーザーから報告を受けた事案の進捗を聞く対話システムを題材に, ユーザーの使用頻度が高い言葉に基づいてユーザーの報告内容を表現する「見出し生成モデル」を使ってシステムが獲得したい事案を端的かつユーザーに合わせた形で表現するシステムを提案する. 本研究では, 過去の履歴データから比較的簡単なタスクとなるデータに絞ったうえでモデルを開発し, ユーザーの特徴に合わせた表現が可能で, Rouge 値 0.7 以上の精度となるモデルを構築した. 生成モデルの出力を使った問いかけが, システムから送付されるメッセージとして違和感がないか, 過去事案を思い出すのに役立つか, ユーザー評価した結果, 4 名中 1 名以上が肯定した問いかけは多かったが, メッセージとして不適切なものもあり, フィルタ等の処理など改善が必要であることが分かった.

**キーワード:** タスク指向対話システム, 話題導入, 要約, 見出し生成

## Topic Introduction Considering User Characteristics for Task-Oriented Dialogue System

MAYUMI MATSUBARA<sup>1,a)</sup> NAOKI ASABA<sup>1</sup> SHOICHI NAITO<sup>1</sup> SHINTARO KAWAMURA<sup>1</sup> SHINYA IGUCHI<sup>1</sup>  
MASAKI NOSE<sup>1</sup> NAOKI OKAZAKI<sup>2</sup>

Received: January 18, 2022, Accepted: April 21, 2023

**Abstract:** In recent years, the development of AI technology such as deep learning has brought more demand to utilize task-oriented dialogue systems in business. In order to get more rational cooperation between the system and users, the system should take their past conversation into account during the task operations. From this point of view, we propose an approach to provide the users with appropriate information consisting of simple and customized phrases by exploiting a headline generation model. The headline generation model rephrases the input text from a user as a phrase based on higher frequently used words by referring to the past dialogue. In this study, we apply the headline generation model to our dialog system which asks our in-house sales staff about the progress of their past-reported items. Since we have selected the data so that the model can avoid complex tasks at this stage of the study, the model performs keeping ROUGE score above 0.7 where the generated phrases can be properly characterized by the users. As a result of the subjective evaluations by the users in terms of unnaturalness of the output messages and effectiveness as a reminder of their tasks, we obtained generally positive feedback. On the other hand, we also found that the proposed system still needs some improvement such as adding filtering process to reduce inappropriate outputs.

**Keywords:** task-oriented dialogue system, topic introduction, summary, headline generation

### 1. はじめに

深層学習をはじめとする各種 AI 技術の発展により, タスク指向対話システムを業務に活用するシーンが増えている. たとえば, よくある質問と回答のデータを学習させたモデルを利用した問い合わせ自動応答対話システムや, タ

<sup>1</sup> (株) リコー  
Ricoh Company., Ltd., Ebina, Kanagawa 243-0460, Japan

<sup>2</sup> 東京工業大学  
Tokyo Institute of Technology, Meguro, Tokyo 152-8550, Japan

<sup>a)</sup> mayumi.nakamura@jp.ricoh.com

スクを遂行する際に必要な情報をユーザから対話形式で引き出し、予約等の業務を行うシステムが存在する。AI技術の進展により、今後さらにシステムとユーザが協調して業務を推進することが想定されるが、現在、業務で活用されている対話システムの多くは、ユーザからシステムに話しかけることによってタスクを実施するものが多い。システムとユーザがより協調して業務を行うためには、システムとユーザが過去に話した内容を踏まえて、ときにはシステム側からユーザに話しかけ、必要な情報を獲得して業務を推進する必要があると考えられる。

そこで、システムがユーザに聞きたい内容があるときに、対話のきっかけをユーザに提供するとともに、ユーザがシステムから問われた内容を思い出しやすいように聞くことができるタスク指向対話システムを提案する。ここで、システムから問われた内容をユーザが思い出やすく、かつ対話のきっかけとして利用できる言葉を見出しと定義する。本研究では、図1のような報告支援対話システムを検討する。本システムのユーザは、同じお客様先を複数回訪問する、かつ、訪問するたびにお客様先で実施した業務を会社に報告する必要がある社員である。本システムの利用フローについて、図1を用いて説明する。まず、ユーザはお客様先で業務を行った後、本システムを用いて自身の業務を報告する(①②)。報告内容は報告のたびにログデータとして蓄積される(③)。ユーザの業務特性として、ユーザは同じお客様先を複数回訪問する。本システムは外部データを参照しながら、ユーザが報告し忘れている可能性がある報告を制御部で検出し、すでに報告済みの内容の見出しを生成するとともに(④)、ユーザから情報を得たい報告について、ユーザの特徴を捉えた要約を用い

て問いかける(⑤)流れで動作する。本システムは報告済みの内容の見出しを付記してユーザに話しかけるため、システムは対話のきっかけを提供するとともに、ユーザが問われた内容を思い出しやすい状態を作ることが可能となる。そのことから、ユーザの応答率の向上や応答に要する時間を短縮しながら、ユーザが報告し忘れた情報を獲得できることが期待される。

本研究では、対話データが十分に蓄積されていない状況を想定し社内ですでに蓄積されているデータを用いて図1④の見出し生成モデルの実現を検討した結果を記述する。2章では、対話履歴の利用ならびにユーザの記憶を引き出しやすくするよう設計された対話システムの先行研究について述べる。3章は、業務特性ならびに問題と、その対応方法を述べる。4章では検証目的と検証手順を述べ、5章で検証結果とまとめ、考察を述べる。

## 2. 関連研究

ユーザの記憶を思い出しやすいするための対話システムとして、高齢者の認知症や孤独感を軽減するために開発された傾聴対話システムがある[1]。そのシステムは、ユーザである高齢者の記憶を思い出しやすいための機能が備わっている。具体的には、ユーザの発話を繰り返したり、問い返したりする機能を有している。しかし繰り返し、問い返しの機能は、ユーザが直前にシステムに伝えた内容に対してのみ行うことができる機能であり、過去の対話の内容を用いて、システム側からユーザに話題として提供することは記述されていない。

過去にユーザと話した内容を対話に用いる事例が記述された論文がある[2]。このシステムは、ユーザが過去に話

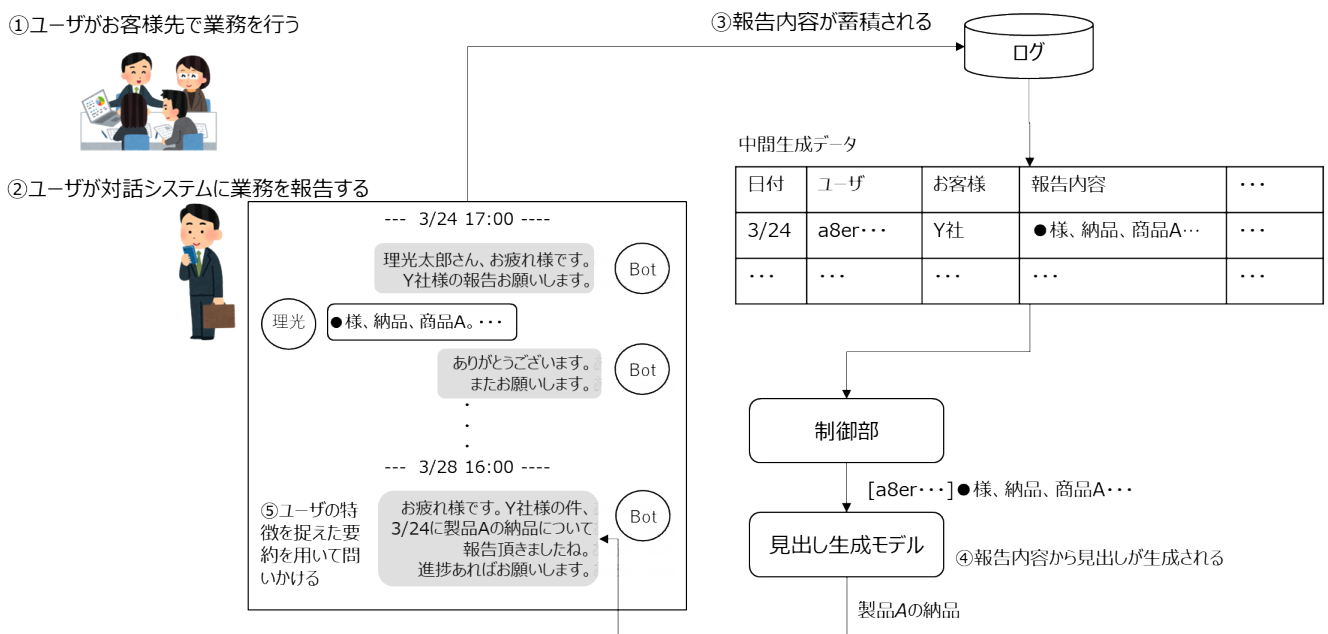


図1 提案する対話のきっかけを提供する対話システム

Fig. 1 Our dialogue system that provides conversation opportunities using a headline generation model.

した対話内容と関係がある発言を再び行った際、対話履歴からユーザ発言の要約を用いて応答するシステムである。しかしながら、過去に話した内容を用いて応答することは可能であるものの、システム側から話題は提示していないので、話題を提供する機能を実現できるものではない。

また、Twitter に投稿されたニュース記事を利用し、対話のきっかけとなる発言を生成する研究が存在する [3]。きっかけとなる発言は記事の要約部分と雑談部分から構成され、対話のきっかけに要約を用いる点で本研究と類似している。しかし、過去のやりとりをユーザに思い出してもらうことが目的ではなく、本研究が着目するユーザの特徴をとらえた要約については検討されていない。他にも、ユーザの興味を引くような要約を生成する研究 [4]-[6] や、システムからユーザに話しかける対話の研究として [7], [8] はすでにあるが、システムに問われた内容をユーザが思い出しやすくするために、ユーザに合わせた要約を対話のきっかけとして用いているものではない。

### 3. 提案内容

#### 3.1 対話システムのタスクおよび想定ユーザ

本システムの想定ユーザは、お客様先を訪問し、営業などの業務を行う社員である。対話システムは、お客様先で業務を行うユーザから業務の報告を受けるタスクを負っている。システムがユーザから報告を受ける事案は1日では完了せず、長期間にわたってユーザが複数回お客様先を訪問し、進捗を確認する必要があるものである。なお、ユーザは1日に複数の事案の報告を行っているため、ときに報告を行うことを忘れてしまう可能性があるため、システム側からユーザに事案の報告を確認する機能が必要と考えられる。

#### 3.2 システム構築に向けての課題

第1の課題として、システムがユーザに事案の進捗を聞く際、多くの事案の対応を同時並行で行うユーザは、いつ実施した何の事案についてシステムから問われているのかを思い出すことが困難になると想定される。そのため、システムがユーザに聞きたい事案を端的に表現したうえで、対話のきっかけをシステム側から提供する必要がある。事案を端的に表現する方法の1つとして、前述で定義した見出しを生成する方法が考えられる。

課題1で述べたような事案の見出しを作成する手段として、ユーザが報告した活動内容を入力とし、見出しが出力されるモデルを構築できれば実現可能となる。そのようなモデルの構築のために、社内にすでに蓄積されていたデータを活用する方法を検討した。具体的には、ユーザが作成した活動予定と活動結果の文書データである。この文書データは、業務特性上、活動予定を最初に入力し、その後活動結果が作成される順序で構成されている。今回作成

したいモデルは活動結果からそれを一言で表現する言葉を作る必要があるため、データが作成される順序と逆の使い方をする必要があり、社内に蓄積されているデータをそのまま使用すると、ユーザが予定を立てた段階では実行するつもりであったことが、結果として行わないケースも存在するため、見出しを生成することが不可能なデータも混在する状況となる(第2の課題)。目的とするモデルの構築には、活動結果に記載されている内容から生成が可能な活動予定のデータを抽出する手段が必要となる。なお、活動予定のデータを見出しに利用することも可能であるが、現実には予定のデータがない活動も存在するため、見出しを生成する手段を検討した。

また、第3の課題として、事案を端的に表現する方法は、ユーザごとに異なる可能性がある。システムがユーザに聞きたい事案を聞くときに、たとえば、業務を行ったときに、会った人の名前を用いた表現のほうが伝わりやすいユーザもいる一方、そのとき自分が実施した内容を含む表現のほうが伝わりやすいユーザも存在すると考えられる。そのため、事案を端的に表現する方法がユーザごとに異なる場合は、ユーザの特徴を捉えた見出しを生成する手段が必要となる。また、第4の課題として、システムからユーザに提示される、事案を端的に表現したものを含む問い合わせは、事案の進捗を聞く問い合わせとしてユーザに受け入れられるかという課題もある。

#### 3.3 提案するシステム

本研究は、対話データが十分に用意できない状況において、ユーザへの問い合わせに利用する見出しを生成する方法について提案する。具体的には、社内に蓄積されている活動予定と活動結果のデータを活用し、見出し生成の問題を活動結果から活動予定を翻訳する問題と捉え、機械翻訳の分野で一般的に用いられている Sequence-to-Sequence [9] (以下 Seq2Seq) のモデルを用いて実現する方法を提案する。Seq2Seq は機械翻訳や要約の分野で用いられる深層学習モデルの1つで、単語や画像の特徴量などの系列を別の系列に変換するモデルである。今回のケースでは、図2

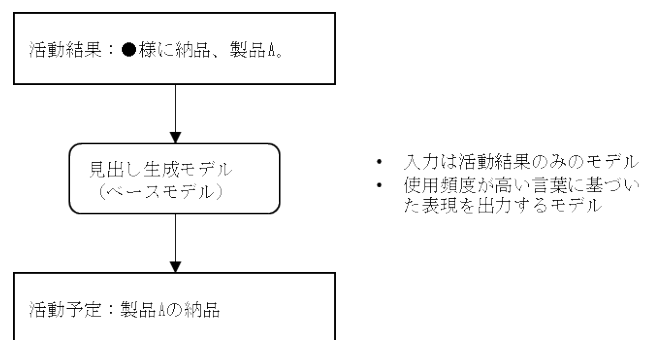


図2 ユーザの特徴を加味する前の見出し生成モデル

Fig. 2 A simple headline generation model.

表 1 学習データの選定

Table 1 Selection of the training data.

#	ユーザ	活動予定	活動結果
1	ab1cd3...	製品 A の納品	●様に納品。製品 A。
2	xb2mr....	B と C の運搬	B を運搬した。

のように、活動結果について記述された自然文を、活動予定について記述された自然文に変換する形で用いる。今回は、Seq2Seq の中でも高い精度が導出可能とされている Transformer [10] を用いてモデルを構築した。活動結果の文字数は、活動予定の文字数よりも多い業務特性がある。そのため、活動結果から活動予定を出力する生成モデルの構築は課題 1 で述べた事案を端的に表現する手段として有効と考えられる。

課題 2 に記載の活動予定と活動結果のデータは、活動予定が先に記述され活動結果がその後に記述される業務特性を考慮する必要がある。これに対して、含意関係に基づく見出し生成タスクの見直し [11] に記載されている方法を参考に、活動予定に含まれる名詞がすべて活動結果に含まれているデータのみ抽出したうえでモデルを構築することとした。たとえば、活動予定と活動結果のデータについて、表 1 の #1 は活動予定に含まれる名詞は「製品 A」と「納品」であり、活動結果に含まれる名詞は「●」「納品」「製品 A」である。活動予定に含まれる「製品 A」「納品」ともに活動結果に含まれているため本データはモデルの構築に利用するデータとした。一方 #2 は活動予定に含まれる名詞は「B」と「C」と「運搬」であり、活動結果に含まれる名詞は「B」「運搬」である。活動予定に含まれる名詞「C」が活動結果には含まれないため、活動結果のデータから活動予定のデータを生成することは不可能である。すなわち、活動予定に含まれる名詞がすべて活動結果に含まれている状態ではないため、モデルを作成するデータからは除外した。

今回構築するモデルは単語の特徴量の系列を別の系列に変換するモデルを用いている。課題 3 に記載のユーザの特徴を考慮した見出しを生成するため、図 3 のように、入力データである活動結果のテキストに加えて、ユーザ個人を識別する情報（以下タグと表現する）をモデルの入力に加えた見出し生成モデルも作成した。ユーザ個人を識別するタグをモデルの入力に用いることで、ユーザの特徴（見出しの書き方など）をモデルが学習することが期待できる。ユーザの言い方に合わせた表現ができればユーザの記憶を想起することができることを期待し、対話システムからユーザに話題を提供する際、スムーズな話題の導入を実現する 1 つの方法であると考えた。

課題 4 への対応として、近年の大規模言語モデルの進展により、実際に対話システムがユーザに話しかけるメッ

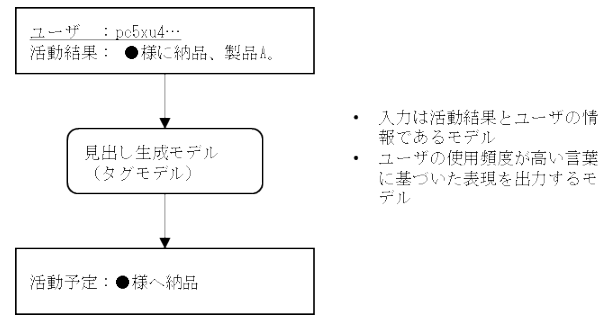


図 3 ユーザの特徴を加味した見出し生成モデル

Fig. 3 Our headline generation model that takes users' taste of word choice into account.

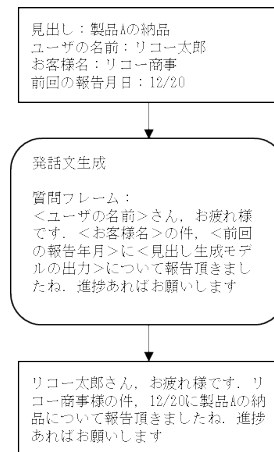


図 4 質問フレームによる問いかけの生成

Fig. 4 An example to generate a friendly reminder by using our message template.

セージをすべて自動生成する方法も可能ではあるが、今回は、会社での業務で活用する対話システムであることを考慮し、安定したメッセージの提供が重要であるため、図 4 のようなルールベースを採用した。

具体的には、対話システムの問いかけは、あらかじめ用意された質問フレームに、見出し生成モデルの出力結果を代入することで生成した。

なお、質問フレームは、「<ユーザの名前>さん、お疲れ様です。<お客様名>の件、<前回の報告年月>に<見出し生成モデルの出力>について報告いただきましたね。進捗あればお願いします。」とした。

## 4. 実験

### 4.1 見出し生成モデルの検証目的

社内にすでに存在する活動予定と活動結果のデータを用い、ユーザの報告内容を端的に一言で表現する見出し生成モデルを構築が可能かを検証することが目的である。検証の観点としては、社内ですでに存在するデータを用いて、ユーザが報告した内容を一言で表せるモデルが構築可能かを検証すること（目的①）、ユーザの特徴を踏まえた見出

しを生成できるかを検証すること（目的②）である。また、ベースモデルまたはタグモデルで生成した見出しを用いた対話システムの問いかけが、ユーザに受け入れられるかを判断することである（目的③）。今回、ユーザへの問いかけを作成する方法として、ベースモデルとタグモデルを比較する形で検証するが、このベースモデル以外のものをタグモデルと比較することも考えられる。たとえば過去の活動結果をそのまま使う、活動結果から抽出した名詞をそのまま出力する、テンプレートに当てはめて出力するなどの方法がある。過去の活動結果をそのまま使う方法について、今回は、本システムの想定シーンに沿わなかったため比較検証しなかった。本システムは、外出先にいるユーザがスマートフォンを通してシステムと対話することを想定したものであるため、スマートフォン画面をスクロール操作することなく読める長さのメッセージが望ましい状況があった。仮に、PC等のデバイスで操作することを想定した場合はこの方法を検討してもよいと考えられる。次に、活動結果から抽出した名詞をそのまま出力する方法について、本システムは将来的に音声対応等の拡張を予定しているため、入出力がテキストから音声が変わった場合でも利用しやすいメッセージを生成したかった状況がある。仮に、名詞が列挙されている箇所を含む問いかけを音声出力する場合、対話としては不自然となる可能性が極めて高いため比較検証しなかった。また、テンプレートに当てはめて出力する方法については、テンプレートやルールの作成および管理運用のコストがかかる懸念があったため実施しなかった。

#### 4.2 方法

実験は、データ準備、データ前処理、モデル構築、評価の順で実施した。

データ準備について、社内に蓄積されている営業日報データの活動予定と活動結果のテキストと、報告を作成した個人を識別することができるIDのデータを用意した。

データ前処理では、スクリーニングと、IDの処理と、データを分ける処理を実施した。はじめに、データスクリーニングについて、この処理は主に次に述べる2つの観点で実施した。ひとつはシステムの対象とした業務の報告に絞るために文字数をつかってスクリーニングした。もう1つは、業務特性を考慮するため、3.3節の課題2への対応に記載した方法でスクリーニングした。具体的には、活動結果と活動予定をともに分かち書きし、活動予定のテキストに含まれる名詞のうち、活動結果テキストに含まれない名詞があるデータを除外した。テキストに記載された名詞の特定は、自然言語処理で一般に利用されているMeCab（辞書はIPADIC）を利用した。次に、IDの処理は、同一個人の手がかりが100件以上存在するユーザには個人を識別することができるタグを、それ以外のユーザには

Other タグを付与し、約400種類強のタグを用意した。最後に、データを分ける処理については、学習、パラメータチューニング、評価の用途として6:2:2になるよう3つにデータを分割した。最終的に学習に利用したデータは約17万件で、活動結果は平均45文字、活動予定は平均11文字のデータで、表2のようなイメージのデータである。たとえば表2の3行目は、活動結果が「●様に納品した。製品Dについては完了。」、活動予定が「製品Dの納品」のデータである。本システムは活動結果から、活動予定に相当する見出しを生成することを検討しているが、この例の場合、活動結果に含まれる名詞「製品D」と「納品」を選び、さらに対話のきっかけとして利用するため、日本語として適切となるよう、助詞を補いながら名詞の語順を入れ替えて見出しを作る必要がある。見出しを作る方法として、単に名詞を抽出するという方法もあるが、これでは対話のきっかけとして利用するには不自然な問いかけになってしまうため、今回見出しを生成する方法を検討した。さらに、この例以外にも、単なる名詞の抽出では、異なる意味として認識されてしまう懸念があるものも存在したため、見出しを生成する方法について検討した。

ここから、モデル構築について述べる。活動結果の単語列から活動予定の単語列を出力するために、今回は深層学習モデルの1つであるTransformerを用いてモデルを構築した。Transformerは系列を別の系列に変換することが可能なモデルで、入力された系列の中から、生成の際、どの単語に注目するのかを考慮することが可能な注意機構を備えた要約の分野でよく用いられているモデルである。Transformerは注意機構を複数個並列で処理することが可能なため処理速度早く精度も高いと言われているため採用した。実行環境は、Ubuntu16.04.6LTS (Xenial Xerus)を利用した。GPUはGeForce RTX2080 Tiで、Python3.8、フレームワークはfairseqを用いた。Transformerの各種パラメータは、Optimizerはadamを、学習率は0.0005、ドロップアウトは0.1、バッチサイズは3584、エポック数は200とした。なお、Transformerの注意機構のヘッドの数は、エンコーダ、デコーダともに8層のモデルを利用した。モデルの学習やパラメータチューニングおよび評価は、前記分割した3種類の各々のデータを用いた。

表2 学習データイメージ

Table 2 An imaginary example of the training data.

タグ	活動結果	活動予定
a1b2...	納品 移動 メンテ ■■ さん に対応依頼	納品 移動 メンテ
x9gu...	××イベント案内 実施	××イベント
G9b8...	●様に納品した。製品 Dについては完了。	製品Dの納品

最後に評価について述べる。まず、目的①に対する評価は、ベースモデルを用いて出力された結果と正解データとの ROUGE-1, ROUGE-2, ROUGE-L の値を算出する形で評価した。ROUGE 値は、要約などの分野で評価指標として一般的に用いられている値であり、1 に近いほどゴールドデータに似ている系列が出力される値である。なお、ROUGE 値の算出は SumEval を利用した。また、各モデルで生成した見出し適切か判断するため、(a) 日本語として不適切な生成となっていないか、(b) 誤解を生む表現なく生成されているかを目視評価した。次に、目的②に対する評価は、前述の目的①のベースモデルに対する評価と同じことを、タグモデルに対して評価する形で実施した。さらに、目的②の評価では、事案を端的に表現する方法がユーザごとに異なるかをデータ解析により確認したうえで、タグモデルの出力がユーザの特徴を満たす形で生成されたか、目視評価も実施した。最後に、目的③に対する評価は、モデルで生成した見出しを用いた対話システムの問いかげが、ユーザに受け入れられるか主観評価した。具体的には、対話システムの問いかげが、システムから送付されるメッセージとして違和感がないか、過去事案を思い出すのに役立つと感じるか否かを評価者に判定してもらった。

## 5. 結果と考察

### 5.1 学習結果

学習の結果、図5のように、Valid データセットにてベースモデルは 141 エポックで、タグモデルは 187 で ROUGE が最大となった。図6のように、Loss の値も確

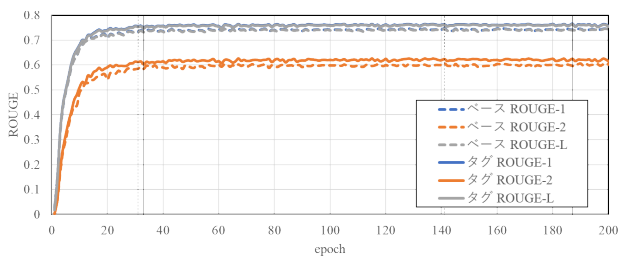


図5 ROUGE 値

Fig. 5 The learning curve in terms of ROUGE score.

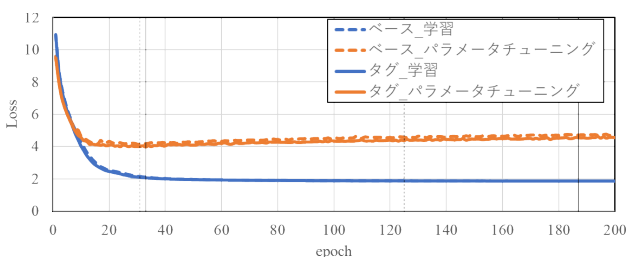


図6 学習過程 (Loss)

Fig. 6 The learning curve in terms of the training loss.

認した所、前述のエポック数で劇的に値が上昇している様子がなかったため今回はこれらをベストモデルと判断した。

### 5.2 評価結果

#### 5.2.1 ROUGE 値

評価データに対する ROUGE 値を表3にまとめた。表では ROUGE-1 を R-1, ROUGE-2 を R-2, ROUGE-L を R-L と表記した。表3の ROUGE 値を確認すると、ベースモデルとタグモデルともに高い値であった。そのためどちらのモデルもユーザが報告した内容を一言で表せていると判断できる。また、ベースモデルとタグモデルを比較すると、タグモデルのほうが約2ポイント高い値となった。また、ベースモデルとタグモデルの ROUGE-1 の値を有意水準 0.05 で検定した結果、有意差が確認された。そのため、ユーザの言い方に合わせた生成ができていると見なせた。なお、今回 ROUGE 値が 0.7 以上と比較的高い数値が出ているが、これは、今回用いたデータが活動結果の平均文字数は 45 字程度、活動予定は平均 11 文字程度のものであったことから、比較的短い文字数のものから短い見出しを生成する簡単なタスクであったことが理由として考えられる。また業務特性を考慮する際に実施したスクリーニングにおいて、活動予定のテキストに含まれる名詞のうち、活動結果テキストに含まれない名詞があるデータを除外したことも、よりタスクを簡単にした要因であると考えられる。

タグモデルは、ユーザの言い方に合わせた生成を実現することを期待して作成したモデルである。そのため、Other タグが付与されたデータはベースモデルとタグモデルとの出力結果に差はなく、個人を識別するタグが付与されたデータに対してはベースモデルよりもタグモデルのほうが ROUGE 値は高くなると予想される。そこで、個人を識別するタグが付与されたデータと、Other タグが付与されたデータ各々に対する ROUGE 値を確認した。その結果、表4のように、Other タグが付与されたデータでは、2つのモデルで差がないが、個人を識別することができるタグが付与されたデータに対する ROUGE 値はベースモデルよりもタグモデルのほうが約3~4ポイント高くなる結果であった。よって、タグモデルは、ユーザの言い方に合わせた生成ができていると考えられる。

表3 ROUGE 値

Table 3 ROUGE scores.

モデル	R-1	R-2	R-L
ベース	0.75	0.61	0.75
タグ	0.77	0.63	0.76
差	0.02	0.02	0.01

表4 ベースモデルとタグモデルの違い

Table 4 Differences between the baseline model and our tag model.

	個人を識別できる タグが付与された データ			Other タグが付与さ れたデータ		
	R-1	R-2	R-L	R-1	R-2	R-L
ベース	0.79	0.66	0.78	0.70	0.54	0.69
タグ	0.82	0.70	0.82	0.70	0.54	0.70
差	0.03	0.04	0.04	0.00	0.00	0.01

### 5.2.2 目視評価

各モデルで生成した見出し適切か判断するため、(a) 日本語として不適切な生成となっていないか、(b) 誤解を生む表現なく生成されているかを目視評価した。具体的には、(a) については、同じ意味の言葉が2回出てきている場合や、人名や商品名などの文字が一部欠落している場合に、日本語として不適切な生成であると判断した。(a) は、日本語として適切かを判断するのではなく、不適切ではないことを確認する形としたのは、今回利用したデータが、ユーザのメモとして業務で使われる場合があることから、必ずしも日本語として適切な状態であることが理想とは言い切れないことが理由である。(b) については、活動結果に記載していない内容が生成された場合は、誤解を生む表現が含まれると判断した。上記判断基準は、主観が介在する領域が少なく客観的に判断できるものであったため、目視評価は筆者自身で実施した。目視評価に用いたデータは、個人を識別するタグが付与できた340名の中からランダムに抽出された35名の評価データ1531件に対する評価を実施した。

その結果、表5のように、ベースモデル、タグモデルともに(a)、(b) いずれの観点でも多くの生成が適切に行われたと判断した。目視評価が高い傾向にあるのは、モデル作成時のスクリーニングを行ったため、比較簡単なタスクにしばらくこまれたためと考えられる。

次に、タグモデルは、システムが獲得したい事案を端的に表現する方法はユーザごとに異なるとの考えのもとに構築したモデルであるが、この考えを確かめるために、前述の検証で用いた35名の評価データ1531件の中から、20名の活動予定と活動結果のデータを20件ずつ計400件抽出し、解析した。その結果、活動結果から特定の内容(たとえば活動の進捗を表す内容など)を活動予定に記載しているユーザが5名存在し、その内容は人によって違いがあることが確認された。また残りの15名も、特定の内容を書く場合と書かない場合があるなど揺らぎはあるが記載内容に傾向はあることが分かった。このことからシステムが獲得したい事案を端的に表現する方法にユーザの特徴があることが確認された。

表5 目視評価結果1

Table 5 Inconsistency evaluation result 1.

モデル	評価データ[件]	a 日本語として 不適切ではない 割合[%]	b 誤解を生む表 現がない 割合[%]
ベース	1531	96	93
タグ	1531	96	94

表6 目視評価結果2

Table 6 Inconsistency evaluation result 2.

モデル	評価データ [件]	ユーザの特徴を満たす件数 [%]
ベース	159	86
タグ	159	94

さらに、この20名の活動結果に対する2つのモデルの出力結果はユーザの特徴を満たす形で生成されているか目視評価した。その結果、表6のように、ベースモデルよりもタグモデルのほうがユーザの特徴を満たす形で生成できている件数が多いことが確認できた。

### 5.3 ユーザ評価

モデルで生成した見出しを用いた対話システムの問いかけが、システムから送付されるメッセージとして違和感がないか、過去事案を思い出すのに役立つと感じるか否かを主観評価した。

評価者には、過去に活動結果に記載の内容を自分が報告したと仮定してもらい、後日、システムから、報告済みの活動結果に関する問いかけを受け取った場合に、その問いかけは、違和感がなく、かつ過去の事案を思い出すことに役立つと感じた場合に1を、そうではない場合に0の評価をつけてもらった。評価者は、客観的な評価ができるよう、本研究に携わっていない社員4名で実施した。評価者は、100件の活動結果に関する問いかけを評価した。問いかけは、ベースモデルを使ったものと、タグモデルを使ったものと2種類存在するため、1名あたり200件、評価をした。評価者がどちらのモデルを使った問いかけか分からないようにするため、評価データの順番をシャッフルした。また、評価の際、判断に迷ってしまうものがあつた場合や活動結果に誤植などを含むものが存在した場合には、評価結果として、0、1以外の値として、対象外の評価も認める形で実施した。

その結果、表7のように、被験者4名全員が、違和感なく過去の報告したシステムに問われた内容を思い出すきっかけになると感じたものはベースモデルで23件、タグモデルで27件存在した。1名以上が肯定した件数も確

表7 システムからのメッセージに対するユーザ評価

Table 7 Functionality evaluation for the output messages from user feedback.

モデル	評価データ [件]	4名が肯定し た件数[件]	1名以上が肯定 した件数[件]
ベース	100	23	92
タグ	100	27	96

認したところ、ベースモデルで92件、タグモデルで96件であった。これは、人によって感じ方に違いがあるからであると考えられる。なお、表7の結果は、肯定の割合が極端に高い評価者が存在する場合に大きく影響を受けるため、各評価者の肯定の割合とその標準偏差を確認した。確認の結果、各評価者の肯定の割合は0.58, 0.66, 0.56, 0.77で、その標準偏差0.09であった。このことから、今回の評価者には極端に肯定の割合が高い評価者は含まれていなかったことが確認できた。

評価者には、肯定しなかった問かけについては、肯定しなかった理由など、コメントを記載してもらった。コメントを解析した結果、日本語として不適切といったコメントが存在した。これは、モデルの出力を質問フレームにそのまま入力したため、たとえば、文の途中で読点が付与されてしまったものや、質問フレームに入力する際、一部同じ言葉が重複して出てきてしまうケースが存在したためである。テンプレートは、「<見出し生成モデルの出力>について報告いただきましたね。進捗あればお願いします。」という部分があるが、モデルの出力の中に「について」という言葉が含まれてしまう場合には、問かけとしては不適切なメッセージになってしまうことがあった。この点は、テンプレート入力前にフィルタ等で処理することで解決可能である。ほかには、活動結果に記載の内容の一部を記載している場合に、より内容を包含し、漏れがない形のほうがよいとのコメントもあった。その他、汎用的すぎる表現で分かりにくい、意味の伝わりやすさや読みにくさを感じたため、といった内容もあった。

#### 5.4 まとめと考察

システムが過去にユーザから報告をうけた事案に対する進捗の情報を獲得するタスクを実行する対話システムを題材に、ユーザの報告内容をユーザが普段よく使用する言葉を用いて表現する見出し生成モデルを用いて、システムが獲得したい事案を端的にかつユーザが思い出しやすい形で提供するシステムを検討した。本研究ではまずはユーザが普段よく使用する言葉を用いて報告内容を端的に表現する機能として、保有済みのデータをもとに作成された見出し生成モデルが利用できそうかどうかを検証した。検証の結果 ROUGE0.7 以上の高い精度を満たす見出し生成モデル(ベースモデル)を作成することができた。また、個人を

識別できるタグを用いたタグモデルでは、ベースモデルよりも高い ROUGE 値となったことから、ユーザの特徴を加味した見出しを生成できることが分かった。目視評価では、日本語として不適切ではないか、誤解を生む表現がないかを確認し、9割強のデータが正しく生成できていることを確認した。なお、今回の検証の値が高い傾向になるのは、研究に用いたデータが、活動結果は平均45文字、活動予定は平均11文字のデータであったことと、業務特性を考慮したスクリーニングを実施したことで、比較的簡単なタスクとなり、活動結果の単語を抽出する形で見出しを生成することが可能となるものが含まれていたことが要因として考えられる。また、質問フレームと見出し生成モデルの出力結果を用いてシステムからユーザに提示する問かけを作成したが、ユーザ評価では評価者全員が肯定と判断したものは少なかったが、1名以上が肯定したものは多く存在することが分かった。ただ、不適切なものも判断されたものは存在し、また、活動結果に記載されている内容を包含する形での生成、汎用的すぎる表現、意味の伝わりやすさや読みにくさに対する改善が必要であることが分かった。

また、本稿では、システムから問われた内容をユーザが思い出しやすくする方法の1つとして、あたかも自分が書いたかのような見出しを生成することを試した。システムから問われた内容を想起させやすくする方法は他にも存在する。たとえば、見出しを書くのが上手い人の情報をタグとして付けて見出しを生成するなどの方法も考えられる。この方法を用いると、ユーザ自身が書いたかのような見出しではなくなるが、より端的に、業務内容を想起させやすい見出しが生成できる可能性がある。本稿では、対話の問かけ部分を生成することについて述べた論文であったが、この問かけ以降もユーザとシステムの対話は続く。問かけ以降の対話については、たとえば定型文で対話を続ける、大規模言語モデル等を用いた対話を行うなど様々な方法がある。今後は、対象となるデータの拡大や、問かけの質を高めるとともに見出しの内容、質の改善などの課題対応が必要である。

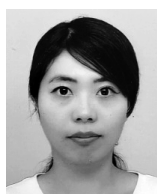
#### 参考文献

- [1] 音声対話ロボットのための傾聴システムの開発: [https://www.jstage.jst.go.jp/article/jnlp/24/1/24\\_3/\\_article/-char/ja/](https://www.jstage.jst.go.jp/article/jnlp/24/1/24_3/_article/-char/ja/)
- [2] 対話システムにおける対話履歴要約の有効性について: <https://cir.nii.ac.jp/crid/1050001339021732480>
- [3] Conversation Initiation by Diverse News Contents Introduction: <https://aclanthology.org/N19-1400/>
- [4] ContextEnhanced Personalized Social Summarization: <https://aclanthology.org/C12-1075.pdf>
- [5] Collaborative Summarization: When Collaborative Filtering Meets Document Summarization?: <https://aclanthology.org/Y09-2005.pdf>
- [6] Enriching Cold Start Personalized Language Model Using



Social Network Information: <https://aclanthology.org/O16-2003.pdf>

- [7] ProDial – An Annotated Proactive Dialogue Act Corpus for Conversational Assistants using Crowdsourcing: <https://aclanthology.org/2022.lrec-1.339/>
- [8] Proactive Human-Machine Conversation with Explicit Conversation Goals: <https://aclanthology.org/P19-1369.pdf>
- [9] Sequence to Sequence Learning with Neural Networks: <https://proceedings.neurips.cc/paper/2014/file/a14ac55a4f27472c5d894ec1c3c743d2-Paper.pdf>
- [10] Attention Is All You Need: <https://arxiv.org/abs/1706.03762>
- [11] 含意関係に基づく見出し生成タスクの見直し: [https://ipsj.ixsq.nii.ac.jp/ej/?action=repository\\_uri&item\\_id=197564&file\\_id=1&file\\_no=1](https://ipsj.ixsq.nii.ac.jp/ej/?action=repository_uri&item_id=197564&file_id=1&file_no=1)



松原 真弓 (非会員)

株式会社リコー 2009年電気通信大学大学院修士課程修了, 同年(株)リコーに入社. 2019年より対話システムの研究開発と業務活用の検討に従事.



麻場 直喜 (非会員)

株式会社リコー 2010年東京農工大学大学院修士課程修了, 同年(株)リコーに入社. 2019年より自然言語処理に関する研究開発に従事.



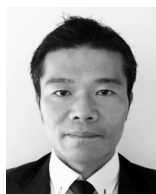
内藤 昭一 (非会員)

株式会社リコー 2010年千葉大学大学院修士課程修了, 同年(株)リコー入社. ソフトウェア開発および自然言語処理の研究開発に従事. 現在, 東北大学大学院博士課程に在籍中.



川村 晋太郎 (正会員)

株式会社リコー 2006年早稲田大学大学院修士課程修了, 同年(株)リコーに入社. LSI回路設計, 無線通信サービス開発を経て, 現在NLPに関する技術探索および研究開発に従事.



井口 慎也 (非会員)

株式会社リコー 1998年神戸大学大学院修士課程修了, 同年(株)日立製作所に入社, 2017年(株)FRONTEO入社, 2020年(株)リコー入社, AIロボット, スマートグラスの研究開発に従事.



能勢 将樹 (非会員)

株式会社リコー 1999年, 富士通(株)に入社し, 主に(株)富士通研究所に在籍. 2013年, (株)リコーに入社. 音声認識やチャットボット, バーチャルエージェントの研究開発に従事しているほか, ディスプレイデバイスの駆動制御や画像処理, 画質評価も専門. 博士(工学).



岡崎 直観 (正会員)

東京工業大学 2007年東京大学大学院情報理工学系研究科博士課程修了. 東京大学大学院情報理工学系研究科・特任研究員, 東北大学大学院情報科学研究科准教授を経て, 2017年8月より東京工業大学情報理工学院教授. 専門は, 自然言語処理, テキストマイニング, 機械学習. 言語処理学会, 人工知能学会, 情報処理学会, ACL各会員.