

小説対話システム *Deep EVE* における LSTM を用いた キャラクター性のある応答文生成

小倉 拓人¹ 谷津 元樹² 原田 実^{2,a)}

受付日 2018年3月8日, 採録日 2018年12月4日

概要: 本研究では, ユーザが指定した小説中の登場人物との自然言語による雑談対話を実現する対話システム *Deep EVE* の開発を行った. 提案システムでは, 青空文庫から取得した小説作品から発話応答の対を抽出しその発話の発話者を意味解析結果に基づきルールベースで推定することで, 発話者情報付きの対話データを自動構築した. この対話データを用いて発話者情報を考慮した対話モデルの学習を行い, ユーザの入力に対し指定された小説作品中の登場人物のキャラクター性を反映した応答文を生成する Seq2Seq モデルを構築した. 応答文の自然さや登場らしさについて, 主観的評価実験を行った結果, 提案システムでは従来システムと比較して, より自然でキャラクター性を反映した応答を生成できることが示唆された.

キーワード: 対話システム, 対話モデル, 機械学習

Response Generation in a Dialogue System *Deep EVE* Using LSTM

TAKUTO OGURA¹ MOTOKI YATSU² MINORU HARADA^{2,a)}

Received: March 8, 2018, Accepted: December 4, 2018

Abstract: We developed a dialogue system *Deep EVE* which responds like a novel character. Consistency of dialogue agent's persona (talking style, background, knowledge and language behavior, etc.) is important to improve the user satisfaction in non-task-oriented dialogue. *Deep EVE* first constructs a dialogue dataset considering speaker name by semantically analyzing novels data. Using this dialogue data, we constructed a Seq2Seq model that considers speakers. This model generates a response considering the speaker's persona vector into which speaker information is embedded. The results of subjective evaluation showed that *Deep EVE* generates better response reflecting characters' personas, and that the naturalness of the response was improved by 20% compared to the conventional system *EVE2014*.

Keywords: dialogue system, dialogue model, machine learning

1. はじめに

対話システムに一貫したキャラクター性を付与することは, 対話の満足度やシステムの娯楽性を向上させるために非常に重要である. 特に雑談のような非タスク指向型対話においては, システムの発話に特定のキャラクター性を与え

ることで, ユーザに対してより親しみやすく良い印象を与えることができる. 近年では, Twitter などから取得した対話データを学習することで, 人手によるルール作成なしに応答生成を実現する機械学習ベースの対話システムの研究がさかんに行われているが, 大規模な対話データをそのまま学習させるだけでは, 口調や語尾, 言葉遣いなどを統一することが困難であり, システムの発話に一貫したキャラクター性を付与することは重要な課題となっている.

キャラクター性を付与した対話システムの試みとして, 白石らは小説作品の登場人物の発話を知識として対話応答を行う用例ベースのシステム *EVE2014* を提案している [1].

¹ 青山学院大学院理工学研究科

Graduate School of Science and Engineering, Aoyama Gakuin University, Sagami-hara, Kanagawa 252-5258, Japan

² 青山学院大学理工学部情報テクノロジー学科

Faculty of Science and Engineering, Department of Integrated Information Technology, Aoyama Gakuin University, Sagami-hara, Kanagawa 252-5258, Japan

a) harada@it.aoyama.ac.jp

EVE2014 では、小説中から発話文を抽出し、その発話者を意味解析結果に基づきルールベースで推定することで登場人物ごとの対話用例を自動構築する。この用例中からユーザ発話と構成する語意や文の構造が類似する発話を検索し、対応する応答を類推することで、登場人物のキャラクター性を持った応答を実現している。しかし、EVE2014 は用例ベースのシステムのため類似発話が存在しない入力に対しては応答を返すことができず、小説から構築した用例は新語や固有名詞などの未知語には対応できないため、ユーザが作品の世界観に沿った入力を行うように配慮する必要があった。また、用例構築の際の話者推定精度も正解率が3割程度と低く、これを改善することも課題としてあげられた。

本研究ではまず、EVE2014 における発話者推定の精度の改善を行い、より正確な対話用例の構築を目指した。次に、小説から自動構築した用例を学習データとして用いて発話者情報を考慮した対話モデルの学習を行い、ユーザが指定した登場人物のキャラクター性を反映した応答を生成する機械学習ベースの対話システム Deep EVE の開発を行った。さらに提案システムでは、EVE2014 では対応できなかった知識のない入力への応答や、機械学習ベースの対話システムの課題である未知語への対応にも取り組んだ。

2. 関連研究

機械学習による応答文生成の代表的なものとして、Sequence to Sequence (Seq2Seq) [2] を活用した手法があげられる。一般的な Seq2Seq では、Encoder と Decoder という2つの Recurrent Neural Network (RNN) を用いる。入力文を1単語ずつ Encoder に入力し、最終的に得られた隠れ層を Decoder 側の隠れ層にわたした後、Decoder に文末記号を入力し、Decoder の出力、すなわち応答文の最初の1単語の生成確率を得る。その後、最も生成確率の高い単語を再帰的に Decoder に入力していくことで、応答文の系列を1単語ずつ生成していく。RNN には、RNN の拡張である Long Short-Term Memory (LSTM) が用いられることが多い。

この枠組みを雑談対話タスクに用いた先行研究として、Vinyals らの Neural Conversational Model [3] や、Shang らの Neural Responding Machine [4] などがあげられる。これらの手法では、学習データとして SNS 上の対話や映画の字幕などから抽出した対話事例を用いており、生成される応答には、入力内容によって「私は医者だ」と答えたり「弁護士です」と答えたりといった、背景知識の矛盾や一貫性のない事例が見られるという課題がある。

機械学習ベースの対話システムにおいてキャラクター性を持った応答生成を行う手法として、Li らは、学習時に発話者情報も同時に学習させる Speaker Model を提案している [5]。Li らの手法では、Twitter やテレビのトランスクリ

プトから構築した話者 ID 付きのデータセットを用い、話者 ID をベクトル表現したものを1時刻前の予測単語と同時に入力することで、その話者の個性を反映した応答を生成しやすくしている。

また、赤間らは大量の Twitter データでモデルを訓練した後、そのパラメータを引き継いで、テレビの字幕などから構築した人物ごとの小規模なデータセットで転移学習を行うことで、その人物の個性を反映したモデルを訓練している [6]。

本研究では、先行研究で多く用いられている Twitter などから対話データを取得する手法ではなく、小説作品から対話を抽出することで学習データの自動構築を目指している。小説作品から対話を取得することの利点としては、Twitter などと比べて絵文字や URL といったノイズがないという点や、日本語の小説であれば、青空文庫などから容易に入手できるという点があげられる。キャラクター性の付与についても、小説の登場人物の個性を反映した応答を学習させることで、各登場人物におけるキャラクター性の違いがイメージしやすくなり、物語中の人物と対話を行いたいという要求を満たすことから、対話システムとしての付加価値の向上にもつながると考えられる。

3. 提案システムの概要

提案システム Deep EVE の概要を図 1 に示す。Deep EVE は主に学習データ自動構築部と応答文生成部で構成される。

学習データ自動構築部では、小説テキストから発話文を抽出し、その発話者を推定することで、発話文とその応答文、さらに応答文の発話者名で構成される学習データを出力する。この学習データを用い、話者を考慮した対話モデルを訓練する。発話者の推定は、小説テキストの各行を意味解析し、その結果に基づきルールベースで発話者を推定する。

応答文生成部では、学習した対話モデルを用いること

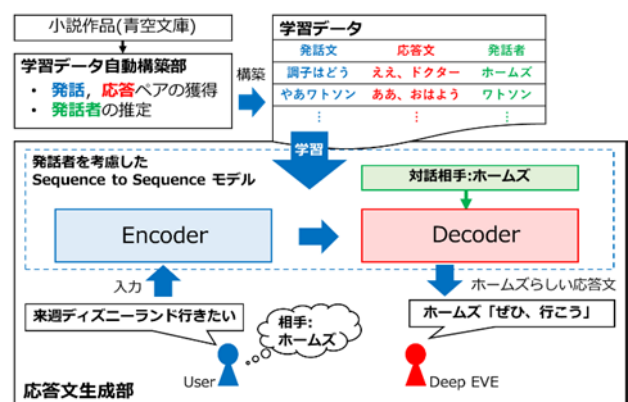


図 1 Deep EVE の概要

Fig. 1 Overview of Deep EVE.

表 1 提案手法の発話者推定ルール
Table 1 Speaker estimation rules.

	発話型	ルール
1	全発話	発話中で人名を名乗っている場合はその人物名を発話者とする。 <ul style="list-style-type: none"> 発話文が一人称の体言に「は」または「が」、および人名を目的語とする断定表現が続く形で構成される場合 「と呼んでくれ」のような発話文 (概念辞書で“願望 3d029d”または“表現する 30f86f”).
2	組込み型	字の文で人名が明示されている場合はその人名を発話者とする。 例: 「〜」とホームズは言った。
3	全発話	前が発話文で、その発話中で人名を呼びかけられている場合、その人名を発話者とする。 例: 「よろしくお願いします、ホームズさん」
4	組込み型	発話の 1 文前→2 文前→1 文後→2 文後の順で探索し、最初に見つかった人名を発話者とする。
5	全発話	連続する発話において、前の前の発話の発話者と推定された人名を発話者とする。

で、ユーザの入力発話に対して指定した登場人物のキャラクター性を考慮した応答文を生成する。対話モデルには、Liらの提案した Speaker Model [7] を改良したものをを用いた。ユーザは入力の際、発話文だけでなく、どの登場人物と対話を行うかを指定することで、システムは入力発話と指定された登場人物情報から、その登場人物の個性を反映した応答文の系列を 1 単語ずつ生成する。この際、入力発話に未知語が含まれていた場合、意味解析結果に基づき小説語彙内の類語に置換することで対応を行う。

話者推定や応答文生成時の未知語の処理などにおける意味解析には、意味解析システム SAGE [7] を用いている。SAGE は、EDR 電子化辞書 [8] の情報を基に、日本語文章の意味解析を行う自然言語処理システムである。このシステムでは、日本語による自然言語文を解析対象とし、前処理として形態素解析と係り受け解析を行った入力を意味解析することにより、語の語意と語間の深層格を含む意味フレームを出力する。

4. 学習データ自動構築

4.1 概要

学習データ自動構築部では、小説作品をテキストとして読み込み、各行を意味解析システム SAGE によって意味解析することで、発話の抽出とその発話者の推定を行う。発話は、作中のテキストのうち“[”と“]”で囲まれた部分とした。発話者推定では、小説のテキストを行単位で読み込み、発話の含まれる行について、その発話の発話者を意味解析結果に基づいて設定したルールによって推定する。従来の EVE2014 では、小説を 1 文単位で読み込み、それが発話文の場合とそれ以外の場合に分けてルールを設定していた。しかし、小説では一般的に 1 行の中に複数の登場人物の発話が登場することは稀であり、『「こんにちは」とホームズは言った。』のように、行中に発話者が明示されている場合も多いことから、本研究では行単位で小説を解析

し、ルールに基づいて発話者を推定することとした。

発話者推定のルールの設定は、2016 年に小井出らにより提案されたルールベースの発話者推定手法 [9] を参考に行った。小井出らは小説中の発話を抽出し、誰が誰と会話しているかを推定することで、敬語や頻度から登場人物の親しさを推定する手法を提案している。この手法では、小説の発話を台詞と地の文で構成される組込み型発話と、台詞のみで構成される独立型発話の 2 タイプに分類し、各分類に基づいた推定パターンを作成することで、優先度順に発話者の推定を行っている。

本研究では、小井出らの設定した発話者推定パターンを参考に、発話者の推定のルールを作成した。小井出らの手法では形態素解析結果に基づいて発話者名やルールに該当するかの判別を行っていたが、本研究では、意味解析システム SAGE を用いて語意などに基づき判別を行う点で異なっている。具体的な推定手法については 4.2 節で述べる。

発話者推定の後、連続する発話の対を 1 対話とし、発話文と応答文をそれぞれ形態素ごとに分かち書きしたものと、推定された応答文側の発話者を 1 件のデータとすることで学習データの構築を行った。学習データ構築については 4.3 節で詳細を述べる。本研究では、青空文庫で公開されているパブリックドメインの小説作品のうち、日本語新仮名遣いのものを 435 編用いることとした。

4.2 発話者推定

4.2.1 概要

提案手法の発話者推定ルールを表 1 に示す。提案手法では、テキストデータとして読み込んだ小説作品の各行について、まず、地の文、組込み型発話、独立型発話の 3 タイプに分類する。発話文を含む 2 タイプの行については、ルールに基づいて人名を探索し、発話者の推定を行っている。なお、人名は SAGE による意味解析の結果から、EDR 辞書中で人名と判断されたものとした。

最も優先度が高いものとして、登場人物が自身の発話中で自分の名前を名乗っている場合が考えられる。ここでは、「わたしはホームズです」のような、1人称 + (は | が) + 人名で構成される文だけでなく、SAGE を用いた意味解析の結果から、「発話文が名詞節と断定節を含む」というルールを追加した。また、EDR 辞書における概念辞書で、上位概念として「願望」や「表現する」を含む発話文も対象とすることで、先行研究では対応できなかった「～と読んでください」や「～とお呼びください」のような多様な名乗り方に対応できるようにした。

次に、『「～」とホームズは言った』のように組み込み型発話の地の文で人名が明示されている場合は、その発話の発話者であるとした。

また、「よろしくお願ひします、ホームズさん」のように、前の発話文の文末において、名前が呼びかけられている場合については、次の発話はその登場人物による発話であると推定できる。

4 番目の優先度として組み込み型発話の場合に発話の前後を表のような順序で人名探索し、それでも発話者が推定できない発話については、発話順による発話者推定を行う。連続する発話について、発話の前の前の発話の発話者を現在の発話の発話者として推定する。ホームズ、ワトソンと対話が継続していた場合、次の発話者はホームズと推定される。これは、Elson らによる「会話は 1 対 1 で交互に行われる」という原則に基づいている [10]。発話順による推定でも発話者が推定できなかった発話の発話者については発話者名を「未定義」とした。

4.2.2 推定精度の評価

提案手法による発話者推定の精度を表 2 に示す。評価に用いた小説作品は、青空文庫の小説から任意のものを 10 編選択して使用した。

提案手法では、実験に用いた 10 作品のうち、最も正解率の高いもので 85%、最も正解率の低いもので「銀河鉄道の夜」で 45%、平均では 62% だった。これは従来 EVE2014 のもの [1] と比べ高い精度であり、小井出らによる推定手法が平均で 50% [9] だったことと比較してもより正確に発話者を推定できると考えられる。

4.3 学習データ構築

話者推定を行った後の発話は、連続する発話の対を 1 セットとして発話者情報を含めた対話データとして出力される。小説作品中に見られる発話には、驚きを表す「えっ」のような極端に短い発話や、1 つの発話が複数の文で構成され、単語数が 100 以上にもなる長文も含まれる発話が含まれる場合が考えられる。極端な長文は機械学習を行う際にメモリを圧迫するため、発話応答の対のうち、発話の場合は台詞に含まれる最後の文を、応答の場合は台詞に含まれる最初の文を抜き出して使用することとした。たとえ

表 2 発話者推定精度

Table 2 Accuracy of speaker estimation.

作品名	正解率
ボヘミアの醜聞	0.64
曲がれる者	0.73
三枚の学生	0.76
まだらの紐	0.60
赤毛連盟	0.85
銀河鉄道の夜	0.45
サセックスの吸血鬼	0.54
株式仲介人	0.53
瀕死の探偵	0.50
橙の種五粒	0.59
10 作品の平均	0.62

ば、発話文「何かね、まさか—火事か？」に対して、応答文として「いや、依頼人だ。年の若いご婦人が興奮気味に來られたらしく～」というような長文の発話があった場合、1 文目のみを抜き出して「何かね、まさか—火事か？」「いや、依頼人だ」を 1 対の発話応答とする。対話データとしては、この発話応答の対のそれぞれを形態素単位で分かち書きし、さらに応答文側の発話者（この場合はホームズ）を追加したものを区切り文字“:”で区切って 1 件のデータとした。

5. 応答生成モデル

5.1 概要

学習データ自動構築部で構築した学習データを用いて対話モデルを訓練する。本研究では、Li らの Speaker モデルにおける発話者情報を、物語の登場人物に置き換えることで、その人物の個性を反映した応答を生成する対話モデルを訓練する。

対話モデルには、入力発話を形態素単位で分かち書きした系列を One-hot ベクトル表現に変換したものを 1 つずつ入力する。この際、One-hot ベクトルに変換できない未知語については、SAGE を用いて類語への置換を行って対応した。この処理を対象の全小説作品について行い、発話者情報付きの学習データセットを自動構築した。評価実験において構築したデータセットの詳細は 6.1.2 項で説明する。

5.2 対話モデル

Li らの提案するモデルを図 2 に示す。単純な Seq2Seq では、多様な背景知識を持つ対話データを用いて学習を行った場合、背景知識などが矛盾した応答が生成される可能性が考えられる。これに対し、Speaker Model では図 2 のように「あなたの出身は？」のような文を Encoder に入力した後、Decoder で応答文を生成する際に、人物名である“Rob”をベクトルに埋め込み、Decoder に対して入力

表 3 Decoder と Speaker-Decoder の出力平均による効果

Table 3 Effect by averaging the output of decoder and Speaker-Decoder.

入力文	Speaker-Decoder のみ	Decoder と Speaker-Decoder の平均
何か食べに行こうよ	ああ、三時ころ帰ったよ	いいえ、ちっとも知りません
何かいいことありましたか？	ああ、三時ころ帰ったよ	まあ、そうですか
あなたの性格を教えてください	ああ、三時ころ帰ったよ	ああ、いいじゃないか
台風が来てますね	ああ、三時ころ帰ったよ	ああ、三時ころ帰ったよ

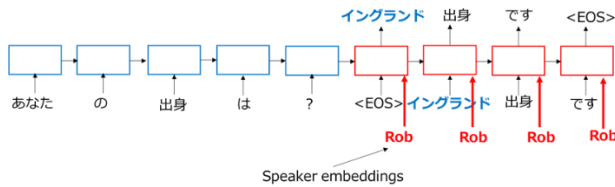


図 2 Speaker モデル (参考文献を基に作成)

Fig. 2 Speaker Model.

文と同時に入力する．ここで、時刻 t における Decoder 側の LSTM 層は以下の式 (1) のように表現することができ、LSTM の内部メモリ c_t を式 (2)、隠れ状態ベクトル h_t を式 (3) によって更新していく．ここで、 i, f, o は LSTM の入力ゲート、忘却ゲート、出力ゲートの各ゲートであり、 σ はシグモイド関数、 W は重みである．また、 x_t は入力された系列のうちの 1 単語であり、 v_i は発話者情報を埋め込んだベクトルである．

$$\begin{bmatrix} i_t \\ f_t \\ o_t \\ l_t \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{bmatrix} W \cdot \begin{bmatrix} h_{t-1} \\ x_t \\ v_i \end{bmatrix} \quad (1)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot l_t \quad (2)$$

$$h_t = o_t \cdot \tanh(c_t) \quad (3)$$

これにより、学習データ中で国や出身地に関連する語として「アメリカ」という単語の頻度が多かった場合でも、「Rob」という発話者情報が考慮され、「Rob」の出身地である「イングランド」という単語が生成されやすくなる．

Li らは、Twitter などから抽出した発話者情報付きの対話データを用いて学習を行っているが、本研究では、小説から構築した学習データを用いている．小説からの発話者情報付きデータを用いた場合の問題として、学習データ内における発話者ごとの発話数の差が考えられる．例として、シリーズをまたいで登場するシャーロック・ホームズについていえば 1,000 発話程度あるが、銀河鉄道の夜など、短、中編作品に 1 度しか登場しない人物については、数発話から 100 発話程度のもも考えられる．そのような登場人物を発話者として指定した場合、キャラクター性を維持した自然な対話を行うことが困難になると考えられる．

データを話者ごとにクラスタリングすることで、学習データが少なくなる問題の対処として、本研究で提案する

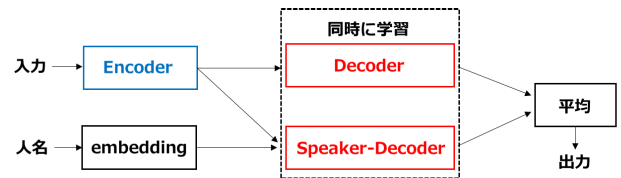


図 3 Deep EVE のモデル構造

Fig. 3 Model structure of Deep EVE.

モデルの構造を図 3 に示す．Deep EVE では、発話者を考慮した Speaker Model における Decoder と、発話者を考慮せず、発話データのみを入力する通常の Decoder の 2 つを同時に訓練し、出力される応答文の各語の生成確率の平均をとることで応答文の生成を行う．

発話数の少ない登場人物（カムパネルラ）を指定した際の、Speaker-Decoder のみを用いた場合の応答と、2 つの Decoder の平均を取った場合の応答の比較を表 3 に示す．

小説から自動構築した学習データを用いて Speaker Model を学習し、発話数の少ない登場人物を指定した場合の応答では、4 つの入力に対してまったく同じ応答を返している．発話者埋め込みベクトルの次元数は学習データ数にかかわらず一定のため、データ数の少ない場合は発話者の情報の影響が大きく応答の多様性が少なくなると考えられる．提案システムでは、話者を考慮しない（したがって学習データの多い）Decoder との平均をとることで、応答の多様性を確保している．

5.3 未知語への対応

機械学習ベースの対話システムでは、あらかじめ設定した語彙以外の語についてはモデルに入力することができないという未知語の問題があげられる．扱える語彙数については多くの研究で 5 万～10 万語程度となっており、それ以外の語については出現頻度に応じて <unk> などのシンボルに置換される．この場合、置き換えられる前の語の意味は考慮されない．小説から構築した学習データでは、新語や特定の固有名詞が出現しないことが考えられるため、小説を基に構築したシステムでは、時事的なトピックを扱うことができず、自然な雑談対話の妨げになる可能性がある．そこで本研究では、モデルに発話文を入力する際、前処理として意味解析を用いて単語間の意味類似度を計算し、小説語彙内の類語への置換処理を行うことで小説データでは

カバーできなかった語意についても考慮することとした。単語間の類似度は、EDR 辞書中の深さをを用いて算出する。語 a と語 b の類似度の計算式を式 (4) に示す。

$$Similarity(a, b) = \frac{2(1-r)dc}{2(1-r)dc + (1-r^{da}) + (1-r^{db})} \quad (4)$$

ここで、 da 、 db はそれぞれ語 a 、 b の共通上位概念から語 a 、 b までの深さであり、 dc は共通上位概念の深さである。 r は公比であり、本研究では 0.5 とした。

例として、「青山学院大学」という単語は学習データ中に存在しないが、「学校」という単語と類似度 0.90 で類似しているため、入力の際に「学校」という類語に置換することで元々の語の意味をある程度保持したままモデルに入力することができる。

5.4 応答文生成

提案システム Deep EVE では、発話を入力する際に、対話を行いたい登場人物の名前を指定してから入力を行う。提案システムにより生成される対話事例を表 4 に示す。対話相手としては、コナンドイルの作品に登場する「ホームズ」、宮沢賢治の「銀河鉄道の夜」に登場する「カムパネルラ」、江戸川乱歩の少年探偵団などに登場する「明智小五郎」の 3 人を指定した。

発話者を指定したことで、システムの応答に「シャーロック・ホームズ」「苹果(りんご)」「ハア」「知ってらい」など、指定した話者ごとに口調や扱う単語に差異が見られた。

また、「青山学院大学」などのような、語彙内に存在し

表 4 Deep EVE の対話事例
Table 4 Sample dialogue with Deep EVE.

話者	発話
ユーザ	私の名前は小倉です
ホームズ	僕の名はシャーロック・ホームズ
ユーザ	私は今日青山学院大学で研究をする予定です
ホームズ	では君とすぐさま出向くとしよう
ユーザ	ワトソンは何歳ですか？
ホームズ	そうですとも
ユーザ	好きな食べ物は何ですか
カムパネルラ	その苹果がそこにあります
ユーザ	相模原に来てください
カムパネルラ	鶴はたくさんいますか
ユーザ	つるはいないと思います
カムパネルラ	ぼくも知ってらい
ユーザ	調子はどうですか
明智小五郎	わたしが明智です
ユーザ	私の趣味は読書をすることです
明智小五郎	ハア、そうですか

ない未知語についても、「学校」という語に置換されることで、応答文も関連した「では君とすぐさま出向くとしよう」というものになっている。一方で、システムの応答には「ハア、そうですか」のような聞き手に回るような相槌的な応答が多く、システム側から話題展開を促すような応答は少ないという課題が見られた。また、相槌的な応答が多いため、質問に対する知識の提供は困難であり、また学習データの対話に含まれない「ワトソンは何歳ですか？」のような質問応答には正しく回答できないという課題も見られた。

6. 評価実験

6.1 実験方法

6.1.1 実験設定

提案システムの有効性の検証を目的に、類義語置換を行わないモデルをベースラインとして主観的評価実験を行った。実験では、小説から構築した学習データを用いて提案システムとベースラインモデルをそれぞれ同様のパラメータで学習し、20 個の入力文を各モデルに入力することで得られた応答文について、男女 11 人を被験者として主観的に評価を行ってもらった。評価項目としては、「日本語として破綻していないか」、「応答として自然であるか」、「登場人物らしい応答文か」の 3 項目とし、5 段階のリッカート尺度で評価を行った。また、作品が発表された当時の口語体や記法による表現の古さを不自然と考慮しないように被験者に説明して評価した。

なお、「登場人物らしい応答文か」の項目については、小説作品を未読の読者でも評価を行えるように、対象の登場人物の台詞を任意に 10 個選択した台詞集を参考資料として呈示し、それを参照しながら評価を行ってもらうこととした。対話相手として、学習データ中の登場人物のうち、推定された発話数が多いものの中から表 5 に示す 10 種を選択し、それぞれの登場人物の応答文について評価を行った。

表 5 評価実験に用いた登場人物
Table 5 Characters used in the evaluation.

登場人物名	登場作品 (作者)
ホームズ	緋色の研究他 (コナンドイル)
ワトソン	緋色の研究他 (コナンドイル)
明智小五郎	少年探偵団シリーズ他 (江戸川乱歩)
宗近	虞美人草 (夏目漱石)
蟻田	火星兵团 (海野十三)
ジョバンニ	銀河鉄道の夜 (宮沢賢治)
カムパネルラ	河鉄道の夜 (宮沢賢治)
須永	彼岸過迄 (夏目漱石)
七郎左衛門	夜明け前 (島崎藤村)
支倉	黒死館殺人事件 (小栗虫太郎)

表 6 実験で用いた応答事例 (一部)
Table 6 Response samples used for evaluation.

対話相手	入力発話	提案システムによる応答	ベースライン (類語置換なし)
ホームズ	海に行ったら、サーフィン をしたいです	帽子をかぶって、来たまえ	わかりました
ワトソン	サッカーは好きですか	さっぱりです	してその動物は何だ?
明智	猫飼いたい、	フーム、その考えはおもしろい	御無用にねがいます
宗近	今日は天気が良いですね。	今日は特別煮え切らない	ああ天気で仕合せだ
ジョバンニ	暇になったので遊びに行き たかったのですよ	もっと遊んでおいで	あなたくじら見たことあって
カムパネルラ	今日は雨でうんざりですね	ああ、僕もそう思っているよ	だけど、その方がいいんだ
須永	この夏は花火大会や海にお 出かけになりましたか	そりゃあ実のところ忙しい男なの で	まあ露西亞派でしょうね
支倉	夏に外にでかけるのもいい のですが熱中症や水の事 故なども怖いですね	では、昨夜の事情を聴かせて貰お う	冗談じゃない

6.1.2 データセット

学習データ自動構築に用いる小説作品には、青空文庫で公開されている現代仮名遣いの小説 435 編を用いた。なお、ホームズのようにシリーズで登場する人物についてはすべて同一の人物であると判断した。これらの小説作品を意味解析し、合計で 55,877 組の学習データを得た。このうち 10% をランダムに選んで開発データとして学習を行った。学習はミニバッチ学習で行い、バッチサイズは 128, 単語ベクトルの次元数は 256 次元、各 LSTM の隠れ層を 256 次元とし、パラメータの最適化には Adam [11] を利用した。このパラメータを用い、話者を考慮したモデルと話者を考慮しないモデルの両方を学習した。実験に用いる入力文には、対話破綻検出チャレンジに用いられる対話破綻コーパス [12] に含まれるユーザ発話の中から、任意の発話文を実験者が各 20 個選択して利用することとした。評価に用いた入力発話文と、各システムの応答事例の一部を表 6 に示す。

6.2 実験結果

6.2.1 日本語としての自然さについての評価結果

10 パターンの各登場人物を対話相手として指定した場合のシステムの応答文について、評価項目「日本語として破綻していないか」についての評価結果の平均を表 7 に示す。

日本語としての自然さについては、類語置換を行った提案システムと類語置換を行わない通常の Speaker モデルの双方で平均して 4 以上の評価を得た。

6.2.2 応答文としての適切さについての評価結果

次に、「応答として自然であるか」の評価項目についての評価結果を表 8 に示す。応答文としての自然さについての評価結果では、提案システムでは、平均して 3.41 だった

表 7 評価結果「日本語として破綻していないか」

Table 7 Evaluation of Correctness as Japanese.

対話相手	提案システム	ベースライン
ホームズ	4.60	4.33
ワトソン	4.54	4.58
明智小五郎	4.56	4.29
宗近	4.53	4.48
蟻田	4.57	4.39
ジョバンニ	4.61	4.20
カムパネルラ	4.45	4.43
須永	4.61	4.46
七郎左衛門	4.58	4.31
支倉	4.53	4.09
平均	4.56	4.36

のに対し、類語置換を行わないモデルの場合は 2.85 と低い評価になった。これは、学習に用いた対話データ量が小さいことと、小説から構築しことにより扱い可能な語彙に制限があったことが考えられる。例として、「サッカーが好きですか?」のような入力があった際に、類義語置換を行わないものでは「サッカー」という単語が<unk>に置き換えられてしまい、「してその動物は何だ?」というようなまったく関連性のない応答文を生成してしまうような事例が見られた。実験に用いた青空文庫で公開されている小説作品は、作者の死後 50 年以上経っているものがほとんどであり、抽出される発言も時代があったものが多いため、扱える語彙に偏りがある。そのため、現代の口語的な入力文を用いた本実験では、多くの語が<unk>に置換されてしまい、適切な応答を返すことができなかった。提案システ

表 8 評価結果「応答として自然であるか」
Table 8 Evaluation of Naturality of Response.

対話相手	提案システム	ベースライン
ホームズ	3.53	3.00
ワトソン	3.25	2.50
明智小五郎	3.53	2.77
宗近	3.13	2.96
蟻田	3.26	2.52
ジョバンニ	3.25	2.77
カムパネルラ	3.58	3.05
須永	3.55	3.04
七郎左衛門	3.49	2.95
支倉	3.51	2.90
平均	3.41	2.85

表 9 評価結果「登場人物らしい応答文か」
Table 9 Evaluation of characteristics.

対話相手	提案システム	ベースライン
ホームズ	4.07	3.52
ワトソン	3.86	3.65
明智小五郎	4.15	3.43
宗近	3.78	3.45
蟻田	3.91	3.31
ジョバンニ	3.66	3.51
カムパネルラ	3.97	3.53
須永	4.07	3.69
七郎左衛門	3.68	3.34
支倉	4.21	3.40
平均	3.94	3.48

ムの場合は、小説に登場しない語についても語彙内の類語に置換するため、偏った語彙を用いた場合でもある程度自然な対話を行うことができたと考えられる。

6.2.3 キャラクター性の反映についての評価結果

評価項目「登場人物らしい応答文か」についての評価結果を表 9 に示す。登場人物らしさの項目では、提案システムで平均 3.94 と高く、登場人物のキャラクター性の応答を反映した応答を生成できていると評価された。

評価が高かったものとしては、ホームズの「帽子をかぶって、来たまえ」というように、語尾や態度にキャラクター性が反映されているものや、明智小五郎の「フーム、その考えはおもしろい」に表れる「フーム」のように作品の文体を顕著に反映したのが見られた。そのため、語尾や態度にあまり特徴が見られないワトソンや、発話集だけでは性格や態度までイメージしにくい宗近や七郎左衛門などの一般的に認知度の低い人物については評価が低くなったと考えられる。また、生成される応答には相槌的なものが

多かったため、特定のキーワードを基に登場人物らしさが評価される事例は少なかった。

7. おわりに

本研究では、小説中の登場人物との対話を行う対話システム Deep EVE の開発を行った。小説作品から発話文を抽出し、意味解析結果に基づき発話者を推定することで、発話者情報付き対話データを自動構築した。構築した学習データを用いて話者を考慮した対話モデルを訓練することで、ユーザが指定した小説中の登場人物との対話を行うシステムを構築した。

評価実験の結果、意味解析ベースで類語置換を行うことで、小説から構築した偏りのある語彙を持つ学習データを用いた場合でも、ある程度適切でキャラクター性を反映した応答を行うことができると示唆された。

一方で、システムの応答は相槌的な応答が多く、話題展開を促す応答を生成することが課題としてあげられた。また、具体的な質問応答も困難であり、地の文や文脈などを考慮する必要があることも課題となった。今後は、そうした文脈や地の文を考慮しながら、発話者情報を考慮しながらも、より多様性があり、話題展開を促すような応答を実現することを目指す。また、より大規模なデータを扱った学習や、用例ベースなどの他システムとの併用も検討していく。

参考文献

- [1] 白石 仁, 五十嵐聡, 原田 実: 質問応答と類推応答による物語との対話システム EVE, 情報処理学会研究報告, Vol.2013-NL-211, No.10, pp.1-6 (2013).
- [2] Sutskever, I., Vinyals, O. and Le, Q.V.: Sequence to Sequence learning with neural networks, *Advances in Neural Information Processing Systems*, pp.3104-3112 (2014).
- [3] Vinyals, O. and Le, Q.V.: A Neural Conversational Model, *Proc. International Conference on Machine Learning* (2015).
- [4] Shang, L., Lu, Z. and Li, H.: Neural Responding Machine for Short-Text Conversation, arXiv preprint arXiv:1503.02364 (2015).
- [5] Li, J., Galley, M. and Brockett, C.: A persona-based neural conversation model, *Proc. ACL*, pp.994-1003 (2016).
- [6] 赤間怜奈, 稲田和明, 小林颯介, 佐藤翔多, 乾健太郎: 転移学習を用いた対話応答のスタイル制御, 言語処理学会第 23 回年次大会発表論文集, pp.338-341 (2017).
- [7] 原田 実, 水野高宏: EDR を用いた日本語意味解析システム SAGE, 人工知能学会論文誌, Vol.16, No.1, pp.85-93 (2011).
- [8] (株) 日本語電子化辞書研究所: EDR 電子化辞書仕様説明書 (1995).
- [9] 小井出慎, 古宮嘉那子, 佐々木稔, 新納浩幸: 物語における登場人物の親しさ推定, 言語処理学会第 22 回年次大会発表論文集, pp.985-988 (2016).
- [10] Elson, D.K., Dames, N. and McKeown, K.R.: Extracting social networks from library fiction, *Proc. 48th Annual Meeting of the Association for Computational Linguistics*

tics, pp.138–147 (2010).

- [11] Kingma, D. and Ba, J.: Adam: A method for stochastic optimization, *International Conference on Learning Representations (ICLR)* (2015).
- [12] 船越孝太郎, 東中竜一郎, 稲葉通将, 菅原 朔, 高梨克也, 大塚裕子, 小磯花絵, 坊農真弓: 対話破綻検出チャレンジにおける対話破綻データと破綻検出結果の分析, 言語処理学会第 22 回年次大会発表論文集, pp.433–436 (2016).



小倉 拓人 (学生会員)

青山学院大学院理工学研究科理工学専攻知能情報コース. 理工学部情報テクノロジー学科卒業.



谷津 元樹

青山学院大学理工学部情報テクノロジー学科助教. 北海道大学大学院情報科学研究科メディアネットワーク専攻博士後期課程修了. 博士 (情報科学).



原田 実 (正会員)

青山学院大学理工学部情報テクノロジー学科教授. 東京大学理学部物理学科卒業, 同博士課程修了. 理学博士.