対話システムが話し手役になるための オープンドメイン独話生成

Open Domain Monologue Generation for Speaking-oriented Dialogue Systems

稲葉 通将
Michimasa Inaba

広島市立大学大学院情報科学研究科
Graduate School of Information Sciences, Hiroshima City University
ianba@hiroshima-cu.ac.jp

吉野 友香
Yuka Yoshino

同上
yoshino@cm.info.hiroshima-cu.ac.jp

高橋 健一
Kenichi Takahashi

同上
takahasi@hiroshima-cu.ac.jp

keywords: monologue, non-task-oriented, dialogue system, humor generation

Summary

This paper presents an open domain monologue generation method for non-task-oriented dialogue systems to be able to speak their opinions and impressions as a speaker. To generate monologues, we acquire suitable utterances that contain a given topic from Twitter. Our method determines whether utterances have cohesion or not using the support vector machine and concatenate them in a row. It scores the utterance sequences from the aspect of their humor, unexpectedness and speciality in the given topic. We acquire the utterance sequences that ranks high as monologues. Results of an experiment demonstrate that our method can generate amusing and semantically appropriate monologues.

1. ま え が き

既存の非タスク指向型対話システムにおいて、自らの意見や考えを積極的に述べることで対話を主導し、話を進めていくような対話戦略を採用したものはほとんど知られていない。多くのシステムは聞き手にまわり、ユーザの発話に対して意味が通じるように話を合わせるなど、対話を破綻させないことを主眼とする受動的な対話戦略を採用している。例えば、代表的な非タスク指向型対話システムである ELIZA[Weizenbaum 66] や A.L.I.C.E.[Wallace 09], 国際的なコンペティションである Loebner Prize^{*1}に出場した多くの非タスク指向型対話システムや、Turing Test 2014 においてチューリングテストに合格したと発表された対話システム Eugene Goostman^{*2}もそのような対話戦略を採用している。また、ユーザの話を聞くことを目指した対話システムに関する研究[目黒 12]も行われている。

しかし、人間同士の雑談では、対話を主導する話し手とそれを聞く聞き手が相互に入れ替わりながら会話を進めていくという特徴があり、自然な対話の実現のためには、時として対話システムも話し手となり、自らの意見

や感想を話すことができる能力は重要であると考えられる。そこで本研究では、任意の話題について意見や感想、事実や経験などをまとまった形で述べる「独話」を自動生成する手法を提案する。また、対話システムに対するユーザの利用欲求が高まる要因として、ユーザに対し意外性やおかしみを感じさせることが重要であるという分析結果[宮澤 12]を踏まえ、淡々と感想や意見を述べるだけではない、人から見て楽しい独話を生成することを目指す。

独話の生成にあたり、本研究では Twitter データを用いる。Twitter は話し言葉が主に使われていること、ツイートされる話題が幅広いこと、API が公開されており、データの取得が容易であることなど、様々な利点がある。しかし、Twitter にはスパムツイートや、誤字脱字、意味的・文法的に誤っている文などノイズとなるものも数多く含まれている[Higashinaka 14]。また、代名詞などの参照関係の存在により、過去のツイートの内容に強く依存しているツイートも多いため、対話システムがツイートを未処理のまま利用することは困難である。

そこで本研究では、発話獲得手法[稲葉 14]を用いて Twitter からシステムの応答として利用可能な発話を獲得する。この手法は任意のトピックを入力とし、そのトピックに関する複数の発話を獲得する手法である。システム

*1 <http://www.loebner.net/Prize/loebner-prize.html>

*2 <http://www.princetonai.com/>

の応答として利用可能な発話の獲得精度は 95.8%と報告されている。本手法を用いることで、使用可能な状況の制約が少ない汎用性の高い発話が獲得できる。次に、獲得した発話について、発話間の結束性の判定を行うことで 2 発話以上を順序付け、および連結し、独話を生成する。最後に、生成した独話をユーモア性、意外性、与えられたトピックにおける特殊性の 3 つの観点から順位付けを行い、人から見て楽しい独話を生成する。本研究では、この順位付けにも Twitter データを利用する。

2. 関 連 研 究

本研究のように、Twitter データを用いた非タスク指向型対話エージェントに関する研究は近年活発に行われている。例えば、杉山らはユーザの発話から獲得した係り受け関係を持つ文節ペアと、Twitter から抽出した文節ペアを合成し、応答文を生成する手法を提案している [杉山 15]。また、Nio らは映画の脚本における対話データと Twitter データを用いて、用例ベースの非タスク指向型対話エージェントを構築している [Lasguido 14]。その他、Ritter らによる Twitter 上のツイート・リプライから対話モデルを構築した研究 [Ritter 10] など知られている。Higashinaka らは本研究で用いる発話獲得手法と同じく、ルールベースのフィルタリングに加え、機械学習を用いたフィルタリングを行うことで、Twitter から発話を獲得する手法を提案している [Higashinaka 14]。

一方、意見や感想、事実や経験などをまとめた形で述べる独話の自動生成を目指した研究例は非常に少ない。特定のドメインを対象とした研究としては、電子カルテから患者に病状などを説明するための独話を生成した研究 [Williams 07] が存在する。ただし、これは電子カルテに記載されているデータを人が理解しやすいような自然言語文に変換することを目的としたものである。

人から見て楽しい文を生成する試みとしては、ユーモアの自動生成が挙げられる [Binsted 94, Labutov 12]。例えば、Binsted らが開発した JAPE [Binsted 94] は辞書と生成ルールを用いて「What is the difference between a pretty glove and a silent cat? One is a cute mitten, the other is a mute kitten.」のような韻を踏んだ文の生成を行うことができる。また、JAPE を改良し、使用する辞書やルールを大幅に増やした STANDUP [Ritchie 07] も開発されている。Labutov らは概念辞書とテンプレートを用いた手法を提案している [Labutov 12]。これは、「なぜ () は () にあるのか? なぜなら () は () からだ」といった穴埋め式のテンプレートを用意しておき、概念辞書における単語間の関係を利用し、各スロットに単語を適切に埋めることでユーモアを生成するというものである。この手法では例えば「なぜ病院にコンピュータがあるのか? なぜならコンピュータはウイルスをもって」から「だ」といった文を生成することができる。

国内においても言葉の読みの同一性に着目した駄洒落やなぞなぞ生成に関する研究が行われている [ピンステッド 98, 濱田 08, 金久 11]。そのうち、対話システムに関連した研究としては、吉田らは概念辞書など複数の言語資源を用いて、駄洒落などのユーモアを含む応答を生成する手法を提案している [吉田 14]。しかし、上に挙げた研究は全て辞書を用いた手法であることから、使用可能な語彙が大きく制限されているという欠点がある。したがって、本研究が対象とするような話題が限定されない文の生成は困難である。

その他、一貫性があり、かつ人から見て楽しい文章を生成する研究としては、物語生成の研究がある。物語生成研究としては、星新一のショートショート自動生成を目指す「きまぐれ人工知能プロジェクト 作家ですよ」[松原 13] が開始されるなど、近年活発に研究が進められている。物語生成のアプローチとして代表的なものはプランニングと推論に基づく手法 [Meehan 80] である。これは、与えられた初期状態から特定の目標状態になるような登場人物の行動系列を生成し、その行動系列を自然言語文に変換するというものである。このプランニングに基づく手法では、登場人物が取ることのできる行動の種類が増加すると、組合せ爆発が発生するという問題があり、最近では、モンテカルロ法を用いて登場人物を行動させることで目標状態に達する行動系列を生成する手法 [Kartal 14] も提案されている。しかし物語生成は、自然言語の表層表現と一対一で対応する意味表現を用いてプランニングや推論を行っており、本研究が対象とするような言語処理はほとんど行わない。

文の順序付けという観点からみると、複数文書要約の研究とも関連性がある。複数文書要約では、複数の文書から重要文を抽出することで単一の要約を生成する。この際、文を適切に順序付けすることで要約の可読性が向上することが知られている [Barzilay 02]。多くの研究では抽出元の文書の情報 (文書の作成日時や抽出した文の位置など) を利用して順序付けを行っているが、Lapata は抽出した文の情報のみを用いた確率モデルに基づく順序付け手法 [Lapata 03] を提案している。本論文ではこの手法との比較も実施する。

3. 発話獲得手法

3.1 概 要

本研究で生成する独話は、任意の話題に関する複数の発話文から構成される。以下では、その発話文を獲得するために用いる発話獲得手法 [稲葉 14] について述べる。

本手法は任意のトピックを入力とし、そのトピックを含み、対話システムの発話として適切な文を Twitter データから獲得する手法である。ここで発話として適切な文とは、その 1 文だけで意味が理解可能であり、かつ対話において使用可能な時間的・空間的状況が限定されてい

ない文を意味する。本手法は、発話として不適切な文を厳しい基準を設け積極的に排除することで、高精度な発話の獲得を実現している。

本手法ではまず、入力されたトピックを含む文を Twitter データから抽出する。次に、フィルタリングルールを適用し、発話として不適切な文を除外する。そして、教師データからの学習に基づく文の点数付けを行い、点数がしきい値以上となった文を発話として獲得する。

3.2 Twitter からの文抽出とフィルタリング

本節では、次節で述べる点数付け対象となる文の取得手順について説明する。

最初に、入力されたトピックで Twitter データを検索し、トピックを含むツイート抽出する。次に、各ツイートを句点および記号で文単位に分割する。それらの文のうち、トピックを含み、「@」(ユーザ名)を含まず、単語数が 6 個以上 30 個未満の文を取得する。

次に、取得した文について、以下の i~iii で述べる 3 種類のフィルタリングルールを適用し、除外されなかった文を点数付け対象の文として獲得する。

i. 分割に失敗した文

取得した文は句点および記号でツイートを分割して得たものであるが、Twitter では「…」や「(」などの記号が文の途中で挿入されることも多く、文が正しく分割されない場合がある。そこで、特定の位置に特定の品詞が存在する文を除外する。例えば、通常は文頭に助詞は置かれたいため、文頭に助詞が存在する文は分割に失敗したものとみなし、除外する。

ii. 時間や場所が限定的な文

使用可能な時間的・空間的状況が限定されていない発話の獲得を行うため、時間に関する語、例えば「今日」や「先週」などの語を含む文は除外する。

iii. 意味や意図が 1 文だけでは理解不能な文

使用可能な文脈が限定されていない発話の獲得のため、1 文だけで意味が理解できない文は除外する。そのため、代名詞を含む文は除外する。また、「～よりも」のような比較を含む文において、比較対象(「～の方が」)が含まれない文のように、省略が起こった文は除外する。

3.3 文の点数付け

本手法では、文中の「助詞以外の単語」および「助詞とその前後の語の品詞 3-gram(以下、助詞 3-gram)」の点数を教師データから学習し、その点数を用いて文を点数付けする。

助詞以外の単語の点数には、その単語が教師データ中の正解発話の総単語数に占める割合と不正解発話の総単語数に占める割合の比を用いる。単語 w の点数 x_w は以下の式により計算する。

$$x_w = \frac{freq_{w_{correct}}}{freq_{all_{correct}}} \div \frac{freq_{w_{incorrect}}}{freq_{all_{incorrect}}} \quad (1)$$

式中の $freq_{w_{correct}}$ は教師データにおいて、単語 w の正解発話全体における出現回数、 $freq_{all_{correct}}$ は正解発話全体の総単語数、 $freq_{w_{incorrect}}$ は単語 w の不正解発話全体における出現回数、 $freq_{all_{incorrect}}$ は不正解発話全体の総単語数である。したがって、点数が 0 に近ければ近いほど不正解発話に出現しやすい単語であることを示し、1.0 より大きければ大きいほど正解発話に出現しやすい単語であることを示す。

助詞 3-gram の点数については助詞 3-gram を一つの単語とみなし、単語の点数と同様、式 (1) により計算する。

文の点数 q は、式 (1) により計算された点数を用いて、以下の式により計算する。

$$q = \prod_{w \in W} \tau_{\theta}(x_w) \quad (2)$$

$$\tau_{\theta}(x_w) = \begin{cases} x_w & (x_w < \theta) \\ \theta & (x_w \geq \theta) \end{cases}$$

ここで、 w は文中に含まれる単語および助詞 3-gram を表す。また、 x_w は教師データから計算される w の点数である。 θ は点数の上限を決定するパラメータである。実験では元の文献 [稲葉 14] と同じ $\theta = 1.4$ を用いた。

文の点数は、単語および助詞 3-gram の点数が θ より小さい場合はそのまま掛け合わされ、そうでない場合、 θ が掛け合わされることにより計算される。このように、本手法では掛け合わされる点数に上限を与えることにより、文の点数が下がりやすく上がりにくい点数付けを行うことで、不適切な文を積極的に排除し、高精度な発話獲得を実現している。

最後に、文の点数 q がしきい値以上の文を発話として獲得する。本論文では、しきい値は 1.0 とした。

なお、実験では語尾の統一のため、ルールに基づいて語尾変換処理を行い、その発話を独話生成に用いる。

4. 独話生成手法

4.1 概要

本章では、前章で述べた手法により獲得した発話を用いて独話を生成する手法について述べる。

提案手法では、獲得した発話間で順序付きペアを作り、結束性の有無を判定することで発話を連結し、独話を生成する。次に、独話の構成をユーモア性と意外性、およびトピックにおける特殊性の 3 つの観点からスコアリングすることで順位付けし、上位のものを適切な独話として出力する。

4.2 発話の結束性判定

獲得した発話は入力したトピックを必ず含んでいるものの、それらをただ羅列するだけでは自然な独話にはならない。例えば、あるドラマで生成した発話には、出演している役者について述べているものもあれば、ドラマ

表 1 結束性判定に用いる素性

f_1	単語（名詞・動詞・形容詞）のペア
f_2	2 単語目にトピックを含む 3-gram のペア
f_3	発話末尾の文節のペア
f_4	同一の名詞の存在
f_5	同一の動詞の存在
f_6	同一の形容詞の存在

のストーリーに関するものもある。それらを区別せず連結してしまうと、脈絡がなく一貫性のないものとなる。

そこで、獲得した発話からすべての組み合わせの順序付きの発話ペアを作成し、発話間の結束性を機械学習を用いて判別を行う。発話間の結束性は、2 発話間に意味的なつながりを有し、かつ矛盾していないことと定義する。判別器には Support Vector Machine (SVM) を用いる。SVM のカーネルは線形カーネルとし、発話間の関係を捉えるための素性は表 1 に示した 6 種類を用いる。素性 f_1 では、各発話に含まれる名詞・動詞・形容詞（トピックを除く）を用いて発話間でペアを作り、結束性判定に用いる。素性 f_2 では、まず発話中のトピックをシンボル [TOPIC] に置き換え、発話の最初と最後にシンボル [BOS] と [EOS] を付加する。その上で、[TOPIC] が 2 番目に位置する 3-gram のペアを判定に用いる。例えば、「ココアっておいしいよね」と「ココアはおいしいし、健康にも良いらしいよ」という 2 発話（トピック：ココア）であれば、 f_1 により「おいしい-おいしい」「おいしい-健康」「おいしい-良い」が、 f_2 により「[BOS]/[TOPIC]/って-[BOS]/[TOPIC]/は」が、 f_3 により「おいしいよね-良いらしいよ」がそれぞれ結束性判定のために用いられる。 f_4 と f_5 に関しては、今回の例では発話間で同一の名詞と動詞が存在しないため、使用しない。一方、共通する形容詞「おいしい」が発話間に存在するため、 f_6 により、「同一の形容詞が存在」が判定のために用いられる。なお、素性 $f_1 \sim f_3$ に関しては、学習データ中で 2 回以上出現したもののみ、結束性判定のために用いる。

SVM による判定の結果、結束性を有しないと判定された発話ペアは除外する。

最終的に SVM による判定の結果、結束性を有すると判定された発話ペアを連結することにより、3 発話以上からなる独話（発話の系列）を生成する。発話ペアの 2 発話目と、それとは別の発話ペアの 1 発話目が同一であるものを連結することで、3 発話からなる発話系列（独話）が生成できる。例えば発話を s としたとき、 $\langle s_1, s_2 \rangle$ という発話ペアと $\langle s_2, s_3 \rangle$ という発話ペアからは、 $\langle s_1, s_2, s_3 \rangle$ という独話が生成できる。ただし、独話内に同一発話は含まないものとする。

SVM による結束性判定性能を評価するため、後述する実験で使用するものと同じ学習データを用いて予備実験を行った。10 分割交差検証を実施したところ、精度は 0.85、再現率は 0.33 であった。よって、再現率は低いも

の、比較的高い精度で結束性の判定が可能であることが確認された。本研究では、正しい独話を全て列挙することが目的ではなく、ユーモア性・意外性のある独話を生成することが目的であるため、再現率よりも精度が重要である。また、Twitter 上で広く言及されているトピックであれば、発話獲得手法は多くの場合、1 トピックにつき 100 個以上の発話が獲得可能である^{*3}[稲葉 14]。結束性の判定対象はそれらの順序付きペアであるため、再現率が 0.3 程度であっても、独話を生成するための十分な数の発話ペアを獲得することが可能である。

4.3 独話の順位付け

本節では、人から見て楽しい独話を獲得するため、前節で述べた手法により獲得した発話系列に対し、順位付けを行う手法について述べる。本研究において良い独話とは、「少しずつ話を掘り下げていく展開となっていること」、「話の最後が最もユーモア性・意外性が高いこと（オチがあること）」と仮定し、順位付けを行う。

本手法では、まず独話内の各発話について「ユーモア性」、「意外性」、「与えられたトピックにおける特殊性」の 3 つの観点から点数付けを行う。次に、この点数を用いて良い発話が上位になるように順位付けを行う。

§1 ユーモア性と意外性

独話内の発話のユーモア性と意外性は学習データを用いて評価する。ここで問題となるのは学習データの収集方法である。ユーモア性や意外性のある文を人手で作成するのは極めて高コストな作業である。Web 上からユーモアを含む文をブートストラップ法により収集する手法も提案されている [Mihalcea 05] が、この手法は URL に「joke」や「funny」などの特定の 6 種類の単語を含む Web ページのみを対象としたものであり、獲得できる文の量には限りがある。

そこで本研究では Twitter におけるリプライに着目することでユーモア性・意外性のある文を効率的に収集する。すなわち、「笑える」「意外だ」という意味のリプライが返されたツイートは、少なくとも 1 名はユーモア性・意外性があると判断したツイートと考えられる。そこで、そのようなリプライが返されたツイートを大量に収集し、学習データとして用いる。同時に、ユーモア性と意外性が少ないツイートも同様の手法で収集する。

提案手法では、ユーモア性と意外性を判定するための学習データ収集のため、リプライ表現として表 2 に示す 9 パターンと、48 種類の文頭表現（「それは」「ほんと」「超」など、リプライ表現に前置されても意味の変わらない表現）を用いる。収集を行う際には、まず Twitter をリプライ表現によって検索し、その結果取得できたツイートから以下の条件をすべて満たすもののみを抽出する。

- 宛先のツイートが存在すること（リプライであること）

*3 15 億ツイート程度の Twitter データを用いた場合。

表 2 学習データ収集に用いた表現

収集対象	リプライ表現
ユーモア性	笑える
ユーモア性	ウケる
ユーモア性	爆笑
意外性	マジか
意外性	マジかよ
意外性	マジっすか
意外性	マジですか
非ユーモア性・非意外性	そうですね
非ユーモア性・非意外性	そうだよね

- リプライ表現, もしくはいずれかの文頭表現が付加されたリプライ表現が文頭に存在すること
- 宛先のツイートが取得可能であること

次に, 抽出したリプライの宛先となるツイートを取得する. こうして, 取得したツイートを収集対象の学習データとして獲得する. 以下, 学習データ中におけるユーモア性を持つツイートをユーモアツイート, 意外性を持つツイートを意外ツイート, 非ユーモア性・非意外性のツイートを標準ツイートと呼ぶこととする.

ユーモア性と意外性の計算には, 評価対象の発話に含まれる名詞・動詞・形容詞 (以下, この3種類の品詞の語を単語とする) に着目する. 標準ツイートに比べユーモアツイート, 意外ツイートに出現しやすい単語を多く含む発話には高い点を与え, 標準ツイートに出現しやすい単語を多く含む発話には低い点を与える. そのために, 発話を s としたときのユーモア性 $H(s)$ と意外性 $U(s)$ は以下の式により定義する.

$$H(s) = \sum_{w \in \text{word}(s)} \frac{d_h(w)}{d_n(w)} \quad (3)$$

$$U(s) = \sum_{w \in \text{word}(s)} \frac{d_u(w)}{d_n(w)} \quad (4)$$

式中の $\text{word}(s)$ は発話 s に含まれる単語の集合を意味する. $d_k(w)$ は対象とするツイート集合 k における単語 w の出現するツイートの割合であり, 逆文書頻度 (IDF) の計算式を元にした以下の式により計算される.

$$d_h(w) = \frac{DF_h(w)}{N_h} \quad d_u(w) = \frac{DF_u(w)}{N_u}$$

$$d_n(w) = \frac{DF_n(w)}{N_n}$$

ここで, 1つ目の式における $DF_h(w)$ はユーモアツイート集合において単語 w が出現したツイート数, N_h はユーモアツイート集合におけるツイート数を意味する. 上記の他の式についても同様であり, 式中の添字 u は意外ツイートを, n は標準ツイートをそれぞれ対象とすることを意味する.

§ 2 トピックにおける特殊性

対象トピックとの共起頻度が低い単語を含む発話は, 一般的なことを述べた発話ではなく, トピックに関して踏み込んだ内容や特殊な内容を含んでいると考えられる. また, ユーモアツイート, 意外ツイートに出現しやすい語であっても, 対象トピックにおいての頻出する語である場合, その語が含まれる発話のユーモア性や意外性は低く, 対象トピックに関して一般的なことを述べた発話である可能性が高い. そこで, 対象トピックと共起しやすい語を多く含む発話には低い点数を, 共起しにくい語を多く含む発話には高い点数を与えることで, 発話の対象トピックにおける特殊性を求める.

トピックにおける特殊性の計算のため, Twitter から対象トピックを含むツイートを抽出する. 抽出したツイートを用いて, 発話 s の対象トピックにおける特殊性 $C(s)$ は以下の式により定義する.

$$C(s) = \sum_{w \in \text{word}(s)} IDF(w) \quad (5)$$

ここで, $IDF(w)$ は1ツイートを1文書として計算した場合の単語 w の逆文書頻度であり, 以下の式により計算される.

$$IDF(w) = \log\left(\frac{N_{\text{topic}}}{DF_{\text{topic}}(w)}\right) \quad (6)$$

式中の $DF_{\text{topic}}(w)$ は, 対象トピックを含むツイート集合において単語 w が出現したツイート数, N_{topic} は対象トピックを含むツイート集合におけるツイート数である.

§ 3 独話の順位付け

本研究では, 少しずつ話を掘り下げていく展開となっており, 最後にユーモア性・意外性の高い発話があること (オチがあること) を良い独話であると仮定している. これまでに求めたユーモア性, 意外性は楽しさを, 特殊性は話の深さを反映した値であることから, これらの値が順に大きくなるように発話が並んでおり, かつ最後に位置する発話は値の絶対値が大きいものが望ましい独話であると考えられる.

そこで, 独話の順位付けのため, 式 (3), (4), (5) により求めた $H(s), U(s), C(s)$ を用いて, 以下の式により独話のスコア r を計算する.

$$r = \sum_{i,j:i < j} \log\left(\frac{\alpha H(s_j) + \beta U(s_j) + \gamma C(s_j)}{\alpha H(s_i) + \beta U(s_i) + \gamma C(s_i)}\right) \quad (7)$$

式中の s_i は独話中の i 番目の発話を意味する. また, α, β, γ はユーモア性, 意外性, および特殊性のバランスを制御するためのパラメータである. 後述する実験では $\alpha = 1.8$, $\beta = 1.5$, $\gamma = 1.0$ を用いた. それぞれのパラメータの値は経験的に決定した.

表3 実験で使ったトピック

地震	失恋ショコラティエ	ドラえもん
台風	ふなっしー	ピカチュウ
ニュース	ジバニャン	パズドラ
楽天	YouTube	Facebook
天気予報	妖怪ウォッチ	Amazon
Twitter	ヤフオク	

5. 評価実験

5.1 概要

提案手法の評価のため、生成した独話の評価実験を行った。本実験では、予め収集した Twitter データを用いて、独話が獲得できることを確認する。性能比較のため2種類のベースライン手法との比較実験も行った。

以下では、使用したデータ、および実験設定について述べる。

5.2 使用データ

3章で述べた手法を用いた発話獲得、および4.3節§2で述べたトピックにおける特殊性の計算に使用した Twitter のデータは、2014年1月から2014年12月までに収集した日本語のツイート約12億件である。

結束性判定のための学習データは上記の Twitter データから発話獲得手法により獲得した発話を無作為に2つ組み合わせ、そのペアの結束性を著者の1人が人手で判定することで作成した。作成した学習データにおける正解発話ペア数は1477個、不正解発話ペア数は2817個の計4294個である。発話獲得のためのトピックは判定者が無作為に選んだものであり、トピックごとの発話数も一定ではない。ただし、学習データ作成時に用いたトピックには、次節で述べる性能評価のために使用したトピックと同一の語、および同義語は含まれていない。

4.3節で述べたユーモア性・意外性判定のためのデータは、2013年8月から2015年4月の間に収集した。データ数はそれぞれユーモアツイートが42615件、意外ツイートを95433件、標準ツイートが86680件である。

5.3 実験設定

本実験で生成する独話は3つの発話からなるものとする。トピックは2014年の Google 年間検索キーワードランキング^{*4}に掲載されているキーワードのうち、提案手法により50個以上の独話が生成できた計17個を用いた。表3に実験で使ったトピックを示した。なお、アルファベットの大文字・小文字については区別せずに扱った。

独話の評価は大学生3名が個別に行った。各評価者は、独話の自然さと面白さの2項目に関して、5段階のリッカート尺度で評価した。

評価は各トピックで生成された各5個、計85個の独話について行った。その際、独話中の最初と最後の発話が重複しないものを上位から順に5件ずつ取り出したものに対して評価を行う。これは、表層的・意味的に類似した独話を複数回評価することを防ぐためである。

また、性能比較のため、以下の2つのベースライン手法との比較を行った。

§1 ベースライン手法1：ランダム

一つ目のベースライン手法は、3章で述べた手法で獲得した複数の発話からランダムに3個の発話を抽出し、連結したものを独話とするものである。発話は重複しないように抽出する。

§2 ベースライン手法2：確率的順序付け法

二つ目のベースライン手法として、複数文書要約のために提案された確率モデルに基づく確率的順序付け法 [Lapata 03] を用いる。確率的順序付け法は複数文書要約における文の並び替えのために提案された手法であり、提案手法と同様、2発話間に含まれる素性の関係を利用した手法である。確率的順序付け法では、発話が $s_1 \dots s_n$ の順に並ぶ確率 $P(s_1 \dots s_n)$ を以下のように定義する。

$$P(s_1 \dots s_n) = \prod_{i=1}^{n-1} P(s_{i+1} | s_i)$$

ここで、2発話の順序に関する確率 $P(s_{i+1} | s_i)$ は以下の式により計算する。

$$P(s_{i+1} | s_i) = \prod_{(f_{(i+1,j)}, f_{(i,k)}) \in s_{i+1} \times s_i} P(f_{(i+1,j)} | f_{(i,k)})$$

ここで、 $f_{(i+1,j)}$ は発話 s_{i+1} に関する j 番目の素性であり、 $f_{(i,k)}$ は発話 s_i に関する k 番目の素性を意味する。また、確率 $P(f_{(i+1,j)} | f_{(i,k)})$ は以下の式により計算する。

$$P(f_{(i+1,j)} | f_{(i,k)}) = \frac{\text{freq}(f_{(i+1,j)}, f_{(i,k)})}{\sum_{f_{(i+1,j)}} \text{freq}(f_{(i+1,j)}, f_{(i,k)})}$$

式中の $\text{freq}(f_{(i+1,j)}, f_{(i,k)})$ は学習データにおいて素性 $f_{(i+1,j)}$ と $f_{(i,k)}$ が共起した回数である。言い換えれば、発話 s_i に k 番目の素性が出現し、かつ s_{i+1} に j 番目の素性が出現した回数を意味する。

実験では、確率 $P(s_1 \dots s_n)$ を確率的順序付け法で生成した独話の順位付けのスコアとして用いる。学習データは提案手法における結束性判定で用いる正解発話ペアを用いる。また、確率的順序付け法では性質上、1発話で定義できる素性しか使用できない。したがって、提案手法における結束性判定で用いた素性のうち、表1の f_4, f_5, f_6 は使用できないため、 f_1, f_2, f_3 に相当する素性である、「単語（名詞・動詞・形容詞）」「2単語目にトピックを含む3-gram」「発話末尾の文節」の3種類を用いる。

5.4 結果

評価実験の結果を表4に示す。表より、提案手法によって生成した独話は、自然さ・面白さの両方で2つのベー

*4 <http://www.google.com/trends/2014/>

表 5 提案手法により生成された独話と評価実験の結果例

トピック	自然さ	面白さ	独話
パズドラ	4.33	4.33	パズドラってなかなか難しいよね。パズドラって難しいけど面白いよね。パズドラは面白いゲームだと思うけど普通のボスについている防御力の設定は苦痛でしかない。
ドラえもん	4.33	4.00	ドラえもんってほんと面白いよね。ドラえもんはアニメより原作コミックのほうが面白い。名作と言われるアニメの実写化で唯一の成功例と言ってもいいものはドラえもんだと思うんだ。
アナ雪	3.66	3.33	アナ雪は確かに面白かったよ。アナ雪が面白かったのは王道のストーリー踏んだからだよ。アナ雪が面白すぎて金ローのハリボタがB級に見えるよ。
Facebook	3.0	3.33	Facebookって面白くないよね。Facebookって素敵なサービスだとは思うけど面白くない。PCのスピーカーから変な音がすると思ったらFacebookの未読通知だったよ。
ヤフオク	2.33	2.67	ヤフオクで中古は安いんだね。ヤフオクでWRの中古外装探しても相場高すぎだよ。ヤフオクで縮全のトップ材探してるんだけどちっこいのしかないよね。
Twitter	1.33	2.00	Twitterでも面白いよね。時差ある人とTwitterでなんか面白い。Twitterで顔晒す人じゃナイン寝すよ。

表 4 評価結果

手法	自然さ	面白さ
提案手法	3.07	3.20
ランダム	2.69	2.45
確率的順序付け法	2.73	2.90

スラインよりも高い評価を得た。また、t検定を実施したところ、提案手法は自然さ・面白さの両方について、ランダムおよび確率的順序付け法に対する有意差が認められた ($p < 0.01$)。以上の結果より、提案手法はベースラインに比べ、質の高い独話を生成可能であることが確認できた。

表 5 に提案手法により生成された独話とその評価結果を示す。表より、提案手法によって生成された独話は、順に話を掘り下げていくような構成になっており、また、最後にはユニーク性・意外性の高い発話が配置される傾向にあることがわかる。一方で、表 5 のトピックが「ヤフオク」で生成された独話のように、発話間に関連性が少なく、独話としてのまとまりがない場合や、トピックが「Twitter」の独話のように、1 文では意味・意図が理解できない発話が含まれている場合は、評価が低くなった。

生成された独話は、以下のようにユーザの応答としてそのまま使用することが可能である。

(独話使用例 1)

システム : 好きな漫画はある?

ユーザ : ドラえもんかな。

システム : ドラえもんってほんと面白いよね。ドラえもんはアニメより原作コミックのほうが面白い。名作と言われるアニメの実写化で唯一の成功例と言ってもいいものはドラえもんだと思うんだ。

また、ユーザが聞き手にまわっている状況では、以下の

ように独話を 1 発話ずつ応答として使用することも可能である。

(独話使用例 2)

システム : パズドラってなかなか難しいよね。

ユーザ : そうだね。

システム : (パズドラって) 難しいけど面白いよね。

ユーザ : うんうん。

システム : (パズドラは) 面白いゲームだと思うけど普通のボスについている防御力の設定は苦痛でしかない。

さらに、この例のように 2 発話目以降のトピックが含まれる文節 (括弧で囲まれた文節) を省略することで、より自然な応答となると考えられる。

5.5 低評価独話の要因

実験結果をさらに詳細に分析するため、評価者が低い評価をした独話について調査を行った。

そこでまず、自然さと面白さの少なくとも一方の評価が 3.00 未満の独話を抽出し、失敗要因を分析した。失敗要因は「不適切な発話の存在」「矛盾の存在」「不適切な構成」の 3 種類とした。このうち、「不適切な発話の存在」は独話中に日本語として不適切な発話や、意味や意図の理解できない発話が含まれている場合に判定する。「矛盾の存在」は意味的に矛盾する 2 つ以上の発話が含まれている場合に判定する。「不適切な構成」は上記の 2 点に該当せず、発話間の意味的な関連が小さかったり、話題の展開が不自然である場合に判定する。なお、「不適切な発話の存在」と「矛盾の存在」のみ重複して判定することを許す。これは、不適切な発話や矛盾を含む場合、結果として構成も不適切になる場合が多く、低評価の要因分析のため「不適切な構成」とは区別して扱う必要があるためである。

表 6 低評価の独話の分析

	発話	矛盾	構成	発話数
提案手法	7	1	24	32
ランダム	6	1	58	65
確率的順序付け法	6	5	42	51

表 6 に結果を示した。表中の発話，矛盾，構成は，不適切な発話の存在，矛盾の存在，不適切な構成にそれぞれ対応しており，各項目の数値は独話の件数を意味する。

表より，不適切な発話が含まれている件数は各手法において同程度であることがわかる。それぞれの手法で用いた発話は，同一の Twitter データと手法を用いて獲得したものであることから，これは妥当な結果であるといえる。

矛盾を含む独話は，提案手法とランダムが確率的順序付け法よりも少なくなった。ランダムに関しては関連性の少ない発話が連結されることが原因であると考えられる。矛盾が生じるためにはある程度の意味的な関連の強さが必要であるが，ランダムで生成された独話にはそれがなかったため，矛盾する可能性が低くなったと考えられる。一方，確率的順序付け法は正例のみから学習を行う手法であり，関連のある発話を順序付けすることは可能であるが，矛盾する組み合わせを排除することができなかったため，より多くの矛盾を含む独話が生成されたと考えられる。したがって，この結果から負例も考慮して独話が生成できる提案手法の有用性が示唆された。

不適切な構成を含む独話は提案手法が最も少なく，他の 2 手法と比較した差も大きい。したがって，提案手法は自然な構成の独話を生成可能な手法であることが確認できた。一方で，提案手法であっても不適切な構成を含む独話が全体に占める割合は大きく，より精度の高い独話生成のためには，不適切な構成の独話を減らすことが最も有効であることが明らかとなった。

6. ま と め

本研究では，任意の話題について意見や感想をまとめた形で述べる独話を自動生成する手法を提案した。本手法ではまず，任意のトピックを入力とし，そのトピックに関する適切な発話を Twitter データから獲得する。次に，機械学習により 2 発話間の結束性を判定し，結束性を有する発話同士を連結することで独話を生成する。生成した独話は，独話内に含まれる発話のユーモア性，意外性，および対象トピックにおける特殊性の観点から順位付けを行い，上位に順位付けされた独話を適切な独話として獲得する。

今後は提案手法により生成した独話を用いる対話システムを構築する予定である。ユーザの対話への満足度を高めるには，対話システムは常に話し手となれば良いわけではなく，聞き手にもまわる必要がある。そのために，

どのようなタイミングで話し手，もしくは聞き手となるかをシステムが判断する必要がある，人間同士の対話の分析をもとに，適切な話し手・聞き手の切り替え手法についても検討していく。

◇ 参 考 文 献 ◇

- [Barzilay 02] Barzilay, R. and Elhadad, N.: Inferring strategies for sentence ordering in multidocument news summarization, *Journal of Artificial Intelligence Research*, pp. 35–55 (2002)
- [Binsted 94] Binsted, K. and Ritchie, G.: An implemented model of punning riddles, in *Proceedings of the 12th National Conference on Artificial Intelligence* (1994)
- [ビンステッド 98] ビンステッド キム, 滝澤修: 日本語駄洒落なぞなぞ生成システム “BOKE”, 人工知能学会誌, Vol. 13, No. 6, pp. 920–927 (1998)
- [濱田 08] 濱田真樹, 鬼沢武久: 同音異義語の意味の多様性を構造にもつなぞなぞの生成, *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, Vol. 20, No. 5, pp. 696–708 (2008)
- [Higashinaka 14] Higashinaka, R., Kobayashi, N., Hirano, T., Miyazaki, C., Meguro, T., Makino, T., and Matsuo, Y.: Syntactic filtering and content-based retrieval of Twitter sentences for the generation of system utterances in dialogue systems, in *Proceedings of International Workshop Series on Spoken Dialogue Systems Technology 2014*, pp. 113–123 (2014)
- [稲葉 14] 稲葉通将, 神園彩香, 高橋健一: Twitter を用いた非タスク指向型対話システムのための発話候補文獲得, 人工知能学会論文誌, Vol. 29, No. 1, pp. 21–31 (2014)
- [金久 11] 金久保正明: 一般的な概念辞書を用いたなぞなぞ質問文生成システム, 情報処理学会論文誌, Vol. 52, No. 8, pp. 2495–2506 (2011)
- [Kartal 14] Kartal, B., Koenig, J., and Guy, S. J.: User-driven narrative variation in large story domains using monte carlo tree search, in *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*, pp. 69–76 (2014)
- [Labutov 12] Labutov, I. and Lipson, H.: Humor as circuits in semantic networks, in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2*, pp. 150–155 (2012)
- [Lapata 03] Lapata, M.: Probabilistic text structuring: Experiments with sentence ordering, in *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, pp. 545–552 (2003)
- [Lasguido 14] Lasguido, N., Sakti, S., Neubig, G., Tomoki, T., and Nakamura, S.: Utilizing Human-to-Human Conversation Examples for a Multi Domain Chat-Oriented Dialog System, *IEICE TRANSACTIONS on Information and Systems*, Vol. 97, No. 6, pp. 1497–1505 (2014)
- [松原 13] 松原仁, 佐藤理史, 赤石美奈, 角薫, 迎山和司, 中島秀之, 瀬名秀明, 村井源, 大塚裕子: コンピュータに星新一のようなショートショートを作成させる試み, 2D-1, 人工知能学会全国大会 (2013)
- [Meehan 80] Meehan, J. R.: *The Metanovel: Writing Stories by Computer*, Dissertations-G (1980)
- [目黒 12] 目黒豊美, 東中竜一郎, 堂坂浩二, 南泰浩: 聞き役対話の分析および分析に基づいた対話制御部の構築, 情報処理学会論文誌, Vol. 53, No. 12, pp. 2787–2801 (2012)
- [Mihalcea 05] Mihalcea, R. and Strapparava, C.: Making computers laugh: Investigations in automatic humor recognition, in *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pp. 531–538 (2005)
- [宮澤 12] 宮澤幸希, 常世徹, 榊井祐介, 松尾智信, 菊池英明: 音声対話システムにおける継続欲求の高いインタラクションの要因, 電子情報通信学会論文誌 A, Vol. 95, No. 1, pp. 27–36 (2012)
- [Ritchie 07] Ritchie, G., Manurung, R., Pain, H., Waller, A., Black, R., and O'Mara, D.: A practical application of computational humour, in *Proceedings of the 4th International Joint Conference on Computational Creativity*, pp. 91–98 (2007)
- [Ritter 10] Ritter, A., Cherry, C., and Dolan, B.: Unsupervised mod-

- eling of twitter conversations, in *Proceedings of the North American Chapter of the Association for Computational Linguistics Human Language Technologies*, pp. 172–180 (2010)
- [杉山 15] 杉山弘晃, 日黒豊美, 東中竜一郎, 南泰浩: 任意の話題を持つユーザ発話に対する係り受けと用例を利用した応答文の生成, *人工知能学会論文誌*, Vol. 30, No. 1, pp. 183–194 (2015)
- [Wallace 09] Wallace, R. S.: The anatomy of A.L.I.C.E., in *Parsing the Turing Test*, pp. 181–210, Springer (2009)
- [Weizenbaum 66] Weizenbaum, J.: ELIZA-a computer program for the study of natural language communication between man and machine, *Communications of the ACM*, Vol. 9, No. 1, pp. 36–45 (1966)
- [Williams 07] Williams, S., Piwek, P., and Power, R.: Generating monologue and dialogue to present personalised medical information to patients, in *Proceedings of the 11th European Workshop on Natural Language Generation*, pp. 167–170 (2007)
- [吉田 14] 吉田裕介, 萩原将文: 複数の言語資源を用いたユーモアを含む対話システム, *知能と情報 (日本知能情報ファジィ学会誌)*, Vol. 26, No. 2, pp. 627–636 (2014)

〔担当委員: 伊藤 敏彦〕

2015 年 6 月 2 日 受理

著 者 紹 介

稲葉 通将(正会員)

平成 24 年 名古屋大学大学院情報科学研究科博士後期課程修了, 同年, 広島市立大学大学院情報科学研究科助教, 現在に至る。対話システム, 対話処理に関する研究に従事 (博士 (情報科学)). 電子情報通信学会, 情報処理学会各会員。

吉野 友香

平成 27 年 広島市立大学情報科学部知能工学科卒業。在学中は主に対話処理の研究に従事。

高橋 健一(正会員)

昭和 52 年 名古屋工業大学情報工学科卒業。昭和 54 年 同大学院工学研究科修士課程修了。同年, 名古屋工業大学工学部助手。同大学講師, 助教授を経て, 平成 6 年 広島市立大学情報科学部教授。現在に至る。工学博士。主に機械学習, パターン情報処理の研究に従事。電子情報通信学会, IEEE 各会員。