

雑談システムにおけるバックチャネルを用いた連続発話生成

福田 拓也[†] 若林 啓^{††}

[†] 筑波大学大学院 図書館情報メディア研究科 〒 305-8550 茨城県つくば市春日 1-2

^{††} 筑波大学 図書館情報メディア系 〒 305-8550 茨城県つくば市春日 1-2

E-mail: [†]sl621632@u.tsukuba.ac.jp, ^{††}kwakaba@slis.tsukuba.ac.jp

あらまし 近年、特定のタスク達成を目的とせず、人間との自然な会話の成立を目指した雑談システムに関する研究が活発に行われている。雑談システムにおける応答手法の1つとして Twitter などの膨大な発話候補データから応答を選択する抽出ベースの方法が存在する。人間同士の会話では単に情報だけを伝え合うのではなく、バックチャネルを駆使しコミュニケーションをより円滑なものにさせている。バックチャネルとは聞き手側が行う会話中の短いリアクションである。本論文ではより人間に近い応答のために、ユーザ発話に対して抽出ベースの手法による応答にバックチャネルを組み合わせた一連の応答生成手法を提案する。実験ではユーザ発話に対して生成した連続発話が意味が通っているものかどうかに関して人手による評価を行う。

キーワード 雑談システム, 連続発話, バックチャネル

1. はじめに

近年、特定のタスク達成を目的とせず、人間との自然な会話の成立を目指した雑談システムに関する研究が活発に行われている。雑談システムとは、特定のタスク達成を目的としない、人間との自然な会話の成立を目指した対話システムである。

雑談システムの応答方法には、あらかじめ用意した発話候補データからシステムの応答として適切な文章を選択する抽出ベースのアプローチと、文生成によって直接発話を作り出すアプローチが存在する。文生成のアプローチはまだ課題が多く、非文を生成してしまうというリスクがあるのに対して、抽出ベースのアプローチは自然な文章から成る発話候補を用いるためおかしな文章を生成することがないという利点がある。発話候補データは Twitter データなどを利用することにより低コストで膨大な量のデータの獲得が可能である [1]。このことから本研究では抽出ベースのアプローチについて検討する。

システムの応答について、人間同士の会話は単に情報だけを伝え合うのではなく、バックチャネルを駆使しコミュニケーションをより円滑なものにさせている [2]。本研究では今石 [3] の定義に基づいて、バックチャネルを典型的な相槌パターンを含めた聞き手側のフィードバックを表す用語として用いる。従来はシステムの扱うバックチャネルは典型的なパターンに限定されていたが福田ら [4] は入力文に対して大規模データから多様なバックチャネルを抽出する手法を提案している。

これを含め一般的な雑談システムの応答は、ユーザとシステムが1回の発話のたびにターンを交代して会話を進めていくことを想定しているが、実際の人間同士の会話では片方の話者が連続して発話することも多くみられる。特に聞き役対話としては単一の発話で応答するのではなく、バックチャネルを組み合わせた応答の方がパターンとして顕著であることが報告されている [5]。これにより、単一の発話で応答を行うよりもこのように連続した複数の発話を用いて応答をすることで、よりうまく

相手とコミュニケーションをとることができる考える。

これまでの雑談システムの応答手法では、このような連続発話を行うことを考慮した手法はほとんど提案されていないが、より人間らしい応答を目指す際に重要と考える。ルールベースの雑談システムでは、バックチャネル応答の後に質問を行うといった応答の工夫を行なっているものが提案されているが、あらかじめ用意されたパターンに従った応答になりがちであり、応答の多様性の点で課題がある。本研究では福田ら [4] のバックチャネル応答抽出手法を用いて、ユーザ発話に対して抽出ベースの手法による応答にバックチャネルを組み合わせた一連の応答生成手法を提案する。

2. 関連研究

バックチャネルによる応答を扱った研究では定型的なパターンのみを対象としているものが多く存在する [6] [7]。バックチャネルを用いたシステムの連続発話の生成では、例えばあらかじめ「なるほど」や「うんうん」などの定型的なバックチャネルのみを対象として行うことは可能である。しかし、人間同士の会話で使用されるバックチャネルは多岐にわたるため、定型的なパターンのみでは聞き役のリアクションとして単調さを感じさせてしまうと考える。このため、より多様なバックチャネルを対象とすることが求められる。近年ではより豊富なパターンのバックチャネルを対象とした応答手法が提案されており、Inaba ら [8] はバックチャネルを返答クラスとして定義し、バックチャネル応答をユーザ発話についての分類問題として扱っている。そして、福田ら [4] はバックチャネルのパターンを制限せず、ユーザ発話に対して適切な応答を Twitter データから抽出する手法を提案している。

また、バックチャネルを含んだシステムの連続した応答を行うものとしては Lala ら [9] は複数の応答コンポーネントによる対話システムの構築をしている。手法のなかでは汎用的なバックチャネル応答を行うタイミングの予測し、「うんうん、どこへ

行ったのですか？」などの応答を行なっている。目黒ら [5] は聞き役対話における特徴を分析し、ルールベースによる対話制御システムの構築を行っている。分析では特に聞き役としては相手に話を聞いているという自分の反応を伝え、そして相手に質問をするパターンが特徴的である。

これらの研究では汎用的なバックチャネルのみ使用され、さらにテンプレートやルールベースのため、生成される連続発話文のパターンは乏しい。提案手法では発話行為ごとに発話文を分類していないが共感や同意などの応答を広くバックチャネルとして扱い、バックチャネルに続くシステムの応答を Twitter データから抽出することで聞き役対話における多様な連続発話の生成を目指す。

3. バックチャネル応答の抽出手法

本研究で生成する連続発話はバックチャネル及びバックチャネルから続く発話で構成される。以下では本研究で用いるバックチャネル応答の抽出手法 [4] について述べる。本手法は Twitter データから、ユーザ発話に合わせた多様なバックチャネルを抽出するものである。

3.1 応答候補文の特徴量

応答候補文の集合 S を Twitter データに含まれる Tweet の集合として構成されるものとする。そして、応答候補文の集合 S は、ある Tweet に対する Reply であり、かつ別の Tweet によって Reply されているような Tweet の集合とする。本手法ではユーザ発話文 u に対して、応答候補文 $t \in S$ のバックチャネルらしさに影響を与えると考えられる特徴量を抽出する関数として、以下の 3 つを定義する。

- (1) ユーザ発話文 u と応答候補文 t の繋がり $f_1(t, u)$
- (2) 応答候補文 t の情報量 $f_2(t)$
- (3) 応答候補文 t の発話促進度 $f_3(t)$

3.1.1 ユーザ発話文と応答候補文の繋がり

応答候補文である Tweet t の Reply 先である Tweet を $g(t) = s$ と表す。応答候補文 t は Tweet s に対する Reply であるから、 s とユーザ発話 u が類似しているほど、 u に対する応答としてつじつまのあった応答が期待できる。このため、ユーザ発話文 u と応答候補文 t の繋がり強さ $f_1(t, u)$ をユーザ発話文 u と Tweet s の TF・IDF ベクトルのコサイン類似度とする。

3.1.2 応答候補文の情報量

バックチャネルの特徴として、情報提供などを行う応答と比べて、発話に含まれる情報量が少ないことが考えられる。発話の情報量 $f_2(t)$ は、応答候補文 t を形態素解析した結果得られる単語列 $w_{t1}, \dots, w_{tN}$ に含まれる各単語の IDF の合計値として求める。これを素性として用いることで、情報量が多い発話であればバックチャネルらしくない応答、小さければバックチャネルらしい応答として判別する手がかりしている。

3.1.3 応答候補文の発話促進度

ある発話によって次にどれだけ長い相手の発話が期待できるかを表す指標を発話促進度とする。これは聞き役として、その次に相手の長い発話を期待できるような発話がより良いバックチャネルの特徴と考えられるためである。ここでは発話候補文

ID	Tweet	Reply
1	センター試験に遅刻する夢を見た	冗談言っていないで勉強しろ。そんなだから浪人するんだよ。
2	猫飼いたい！	まじで！？おれも！
3	インフルエンザに罹ってしまった	明日、サッカーの試合じゃなかった？タイミング悪すぎw
4	ペットを飼うならやっぱ犬だよな	別にそんなことないぞ
5	ペットが欲しいよ〜	本当にねーお金がないから絶対無理だけだね



- ✓ 「！」 「？」 「。」 を区切りとして Reply を分割
- ✓ 2 つに分割できる Reply のみ使用し、Reply から分割された文章をそれぞれ Reply1, Reply2 とする

ID	Tweet	Reply1	Reply2
1	センター試験に遅刻する夢を見た	冗談言っていないで勉強しろ。	そんなだから浪人するんだよ。
2	猫飼いたい！	まじで！？	おれも！
3	インフルエンザに罹ってしまった	明日、サッカーの試合じゃなかった？	タイミング悪すぎw
4	ペットにするならやっぱ犬だよな	別にそんなことないぞ	
5	ペットが欲しいよ〜	本当にねーお金がないから絶対無理だけだね	

図 1 トリプルの作成例

$t \in S$ の発話促進度 $f_3(t)$ を、 t と同じ文字列を持つ発話候補文全てに対する Reply の平均文字数とする。これにより発話促進度が大きい発話ほど良いバックチャネルとして、相手が返答に困るような応答の抽出を避けることが期待できる。

3.2 二値分類によるバックチャネル抽出

ユーザ発話文 u が与えられたとき、それぞれの応答候補文 $t \in S$ について特徴量抽出関数 f_1, f_2, f_3 を適用し、得られた 3 つの実数を並べたベクトルを、 t の特徴ベクトル $\mathbf{x}_t$ とする。この特徴ベクトルからロジスティック回帰を用いてバックチャネルらしさを表すスコアを求め、スコアが最も大きい発話候補文を適切なバックチャネルとして出力する。

4. 提案手法

本研究では入力文と、入力文に対するバックチャネルが与えられた際にバックチャネルに続く応答を抽出する手法を提案する。この手法では Twitter データの Tweet と、Tweet に対する Reply を用いる。Reply は、文章の区切りを示す記号を手がかりに、複数の文章で構成されているものが数多く存在する。今回、記号「！」、「。」、「？」が 1 つ以上連続している箇所を区切りとして Reply を分割し、分割された文章が 2 つの Reply のみを用いる。提案手法では Reply を Reply1 と Reply2 の 2 つの文章に分割できるデータのみ扱い、Tweet - Reply1 - Reply2 のトリプルを 1 データとして扱う。このとき Reply2 は Reply1 から続く文章である。トリプルの作成例を図 1 に示す。

4.1 結束度推定モデル

Tweet - Reply1 - Reply2 のトリプルを入力とするモデルを設計する。このモデルでは Tweet と Reply1 のペアと Reply2 との結束度を推定する。今回設計したモデルを図 2 に示す。

モデルには文章をエンコードするため、BiLSTM を用いる。RNN (Recurrent Neural Network) は可変長のシーケンスを扱うためのモデルとして広く利用されている。これは前の時刻の隠れ層と現在時刻の入力ベクトルを使って現在の隠れ層を更新する。このとき、隠れ層には再帰的に過去の情報がすべて反映された重みがフィードバックされることになる。この過去から現

在へと時間ステップが前向きのシーケンスを処理する RNN に対し, BiRNN (Bidirectional RNN) [10] は過去へと遡る後向きのシーケンスの処理をさらに加えたモデルである. 各時間ステップ t での出力は前向きおよび後向きの双方向からの 2 つの出力ベクトルの連結となる ($\mathbf{h}_t = \vec{\mathbf{h}}_t || \overleftarrow{\mathbf{h}}_t$). これにより双方向のコンテキストを考慮することで RNN よりも良い精度が期待される. しかし RNN では履歴情報が長期になってしまう場合, 誤差をうまく伝搬させられず勾配消失と呼ばれる問題が生じてしまう.

LSTM (Long Short-Term Memory) [11] はこの問題を緩和するための RNN ユニットの代表的な拡張である. LSTM の再帰層の計算は, 式 1 から式 6 で定義される. このとき σ はシグモイド関数, $\mathbf{U}_* \in R^{H \times H}$, $\mathbf{W}_* \in R^{H \times E}$ は重み行列 (H : 隠れ層の次元, E : 入力ベクトル $\mathbf{x}_t$ の次元), $\mathbf{b}_* \in R^{H \times 1}$ はバイアス項である.

$$\mathbf{i}_t = \sigma(\mathbf{W}_i \mathbf{x}_t + \mathbf{U}_i \mathbf{h}_{t-1} + \mathbf{b}_i) \quad (1)$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_f \mathbf{x}_t + \mathbf{U}_f \mathbf{h}_{t-1} + \mathbf{b}_f) \quad (2)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_o \mathbf{x}_t + \mathbf{U}_o \mathbf{h}_{t-1} + \mathbf{b}_o) \quad (3)$$

$$\tilde{\mathbf{c}}_t = \tanh(\mathbf{W}_c \mathbf{x}_t + \mathbf{U}_c \mathbf{h}_{t-1} + \mathbf{b}_c) \quad (4)$$

$$\mathbf{c}_t = \mathbf{i}_t * \tilde{\mathbf{c}}_t + \mathbf{f}_t * \mathbf{c}_{t-1} \quad (5)$$

$$\mathbf{h}_t = \mathbf{o}_t * \tanh(\mathbf{c}_t) \quad (6)$$

今回用いるモデルでは, BiLSTM によりエンコードされた Tweet のベクトルと Reply1 のベクトルを連結させる. その後, 全結合層で Reply2 のベクトルと同じサイズに変換し, 出力されるベクトルに Reply1 のベクトルを加算する. これはより Reply1 の影響を強くするためである. この加算したベクトルとエンコードされた Reply2 のベクトルを $\sigma(\frac{\mathbf{h}_{tweet-reply1} \cdot \mathbf{h}_{reply2}}{|\mathbf{h}_{tweet-reply1}| |\mathbf{h}_{reply2}|})$ として評価し, Reply2 が Tweet と Reply1 から続く文章として適切かどうかを分類する. σ はシグモイド関数である.

学習データの例を図 3 に示す. 学習では元の Tweet - Reply1 - Reply2 のトリプルを正例, 元の Tweet - Reply1 に対してランダムに選択した Reply2 を用いて作成した Tweet - Reply1 - Reply2 のトリプルを負例とする. そして正例として 1 を, 負例として 0 を教師ラベルとする. 実際の Twitter データは自然な発話文とみなせるために正例として用いることができるが, 負例として用いることができる不自然な発話文は元々存在しないためである. これは Reply2 の正確性に焦点を当てた負例サンプリングとみなせる.

4.2 連続発話の生成方法

連続発話の生成の概要を図 4 に示す. 学習済みの前述のモデルの Tweet Encoder にはユーザ発話を入力し, Reply1 Encoder にユーザ発話に対するバックチャネルを入力する. このとき, ユーザ発話に対するバックチャネルは前章で述べた抽出手法を適用し, 取得する. そして, Reply2 Encoder にはあらかじめ用意する応答候補文集合から任意の文章を 1 つ選び入力とし, この文章とユーザ発話 - バックチャネルとの結束度を評価する. 評価後, 結束度が高い文章をバックチャネルと連結させ, 最終

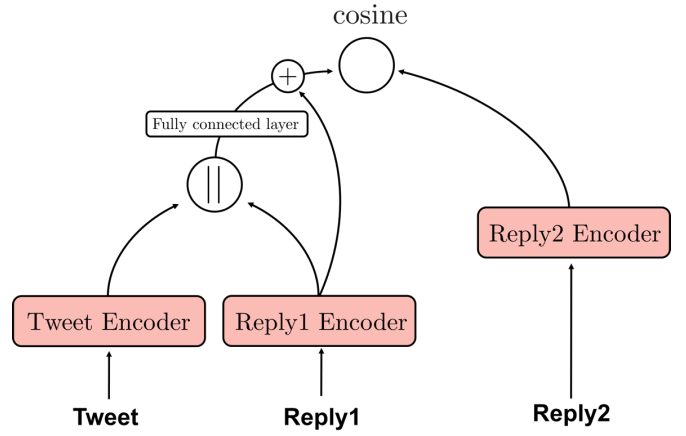


図2 結束度推定モデル

Tweet	Reply1	Reply2	label
センター試験に遅刻する夢を見た	冗談言っていないで勉強しろ。	そんなだから浪人するんだよ。	1
センター試験に遅刻する夢を見た	冗談言っていないで勉強しろ。	タイミング悪すぎw	0
センター試験に遅刻する夢を見た	冗談言っていないで勉強しろ。	種類なんかどうです？	0

図3 学習データの例

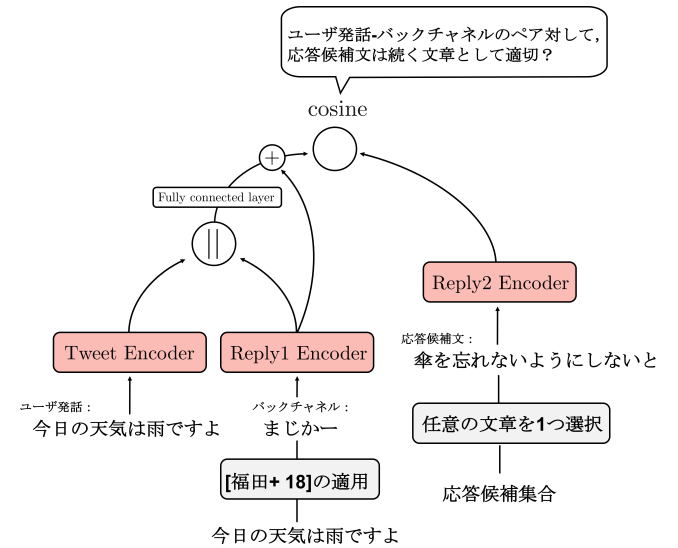


図4 連続発話の生成

的な連続発話として出力する.

例えばユーザ発話として「今日の天気は雨ですよ」という文章が入力されたとする. さらに, ユーザ発話に対するバックチャネルとして「まじかー」という文章が出力されたとする. この「今日の天気は雨ですよ」と「まじかー」のペアと結束度の高い文章を応答候補文の集合の中から選ぶ. このとき「傘を忘れないようにしないと」が最終的に結束度が高い文章と評価されたとする, ユーザ発話である「今日の天気は雨ですよ」に対して, 提案手法より生成される連続発話は「まじかー 傘を忘れないようにしないと」となる.

5. 評価実験

ユーザ発話とバックチャネルのペアを入力し, 提案手法より抽出された文章がバックチャネルから続く応答として適切かど

うか評価する。

5.1 実験用データ

5.1.1 応答候補文の集合と前処理

実験では、2012/04/01～2012/04/07の期間で収集したTwitterデータを使用した。文章のコンテキストがより複雑になってしまったためハッシュタグやURLを含むデータは除いた。また、「。」や「?」などの記号のみの文章であった場合も不適切な文章として除いた。このとき、応答候補文は約1,500万件であった。今回は実行時間の観点から、全ての応答候補文について評価するのではなくTweetが紐づけられたReplyデータを用いて、入力文とTweetの類似度が上位10,000件のReplyを応答候補文として抽出する。提案手法ではバックチャネルに続く応答を、この抽出された応答候補文のサブセットを用いて選択する。

5.1.2 結束度推定モデルの学習データ

学習データは、上記のTweet収集期間には含まれない期間のTweetとReplyからならペアデータを使用した。元のTweet - Reply1 - Reply2のペアを正例とし、Tweet - Reply1に対してランダムにReply2を選択し作成したペアを負例とした。実験では1件の正例に対して2件の負例を作成した。

5.1.3 評価データ

今回は評価データとして、福田ら[4]の評価実験によりバックチャネルとして適切だと評価されたデータを使用する。これは入力文と応答文について、クラウドソーシングにより1データにつき3名に「適切」、「どちらとも言えない」、「不適切」のいずれかで評価してもらった中で、3名全員が「適切」と判断した計273件の入力文とバックチャネルのペアデータである。

5.2 実験方法

実験では提案手法と比較手法のそれぞれ応答文を出力し、クラウドソーシングにより評価する。比較手法はナイーブな手法としてTweetが紐づけられたReplyデータを用いて、入力文とTweetの類似度が最も高い応答候補文を出力するものとする。

評価は被験者に以下のような3つの発話からなる会話を提示し、行う。

Aさん：日本のアニメのクオリティは高いよね！（１）
Bさん：確かに（２）

Bさん：日本には凄腕のアニメーターやイラストレーターが多いね（３）

これは上からユーザ発話、バックチャネル、そしてバックチャネルから続く応答となっており、それぞれ（１）、（２）、（３）と番号がつけられている。また、各発話の話者も示している。最後の応答である（３）が提案手法または比較手法が出力する文章であり、（１）、（２）の文章は各手法で共通である。被験者には（３）の文章が適切かどうかを「○」「▲」「×」で評価してもらう。このとき、○が「（３）は、（１）および（２）ともかみ合っている」、▲が「（３）は、（１）または（２）のどちらか一方とかみ合っている」、×が「（３）は、どちらともかみ合っていない」となっており、○を2点、▲を1点、×を0点とし、その平均値を各手法の評価値とする。

実験はクラウドソーシングサービスのLancers^(注1)を利用し、

表1 バックチャネル応答に続く発話の適切さの平均評価値
(p値<0.001)

手法	平均評価値	標準偏差	件数
提案手法	1.607	0.783	219
比較手法	1.067	0.963	238

1データ（入力文とバックチャネルのペア）につき5件の回答を得る。これにより、各手法について1365件の回答データを用いて評価する。今回、5件の評価のうち同じ評価が3件以上なされたデータのみ使用する。

5.3 結果と考察

各手法の平均評価値を表1に示す。提案手法では3名以上の評価者が同じ評価をしたデータは273件中219件あった。その中で、3名以上の評価者が「（３）は、（１）および（２）ともかみ合っている」としたデータは179件（79.5%）、「（３）は、（１）または（２）のどちらか一方とかみ合っている」としたデータは4件（1.8%）、「（３）は、どちらともかみ合っていない」としたデータは41件（18.7%）となった。提案手法の平均評価値は1.226、標準偏差は0.954であった。また、比較手法では3名以上の評価者が同じ評価をしたデータは273件中238件あった。その中で、3名以上の評価者が「（３）は、（１）および（２）ともかみ合っている」としたデータは119件（50.0%）、「（３）は、（１）または（２）のどちらか一方とかみ合っている」としたデータは16件（6.7%）、「（３）は、どちらともかみ合っていない」としたデータは103件（43.3%）となった。比較手法の平均評価値は1.067、標準偏差は0.992であった。2項検定を行った結果、提案手法と比較手法の平均評価値の差は5%有意水準において有意であった（p値<0.001）。以上より、提案手法の平均評価値は比較手法を上回る結果となり、福田ら[4]と提案手法を合わせて適切な連続発話を生成することができたと考える。提案手法の成功例を表2に示す。

また、図5に提案手法で失敗した全41件のデータについて、出力候補上位30件までを評価してもらい、適切と判断されたデータのヒストグラムを示す。結果は41件中1件のみ適切と評価されたデータが0件であり、それ以外の40件の失敗は全て候補集合に正解はあった。このことから、これらのデータについてはモデルを改良することで適切な応答を抽出し、より高いユーザ評価に近づけられることが期待される。

6. おわりに

本研究ではユーザ発話に対して、バックチャネルとそれに続く応答を組み合わせた連続発話の生成手法を提案した。提案手法ではTwitterデータを活用し、Tweet - Reply1 - Reply2のトリプルを入力とする結束度推定モデルを構築し、バックチャネルの抽出手法と組み合わせて連続発話の生成を行う。実験ではユーザ発話と、ユーザ発話に対するバックチャネルの後に続く応答として適切な文章を抽出できたかを評価した。この結果として、過半数の被験者が同じ評価をしたデータについて提案手法は比較手法よりも有意に高い評価であった。今後の課題としてはモデルの性能向上やバックチャネルの後に続く文章を質問

(注1) : <http://www.lancers.jp>

表2 提案手法の成功例

ユーザ発話	バックチャネル	出力された応答文
お腹の調子がすこぶる悪い。	大丈夫？	お腹大事にしてあげてっ
大学の課題なんもやってないんだよね。そろそろ危機感。	同じくっw	課題ホントやばい w
普通車卒業検定。緊迫した空気が流れております。	がんばれー！	俺応援する！
コミックを買い始めてしまったら全部買わないと気が済まない人もいるんですよ	わかる！笑	そういうのよくあるよね

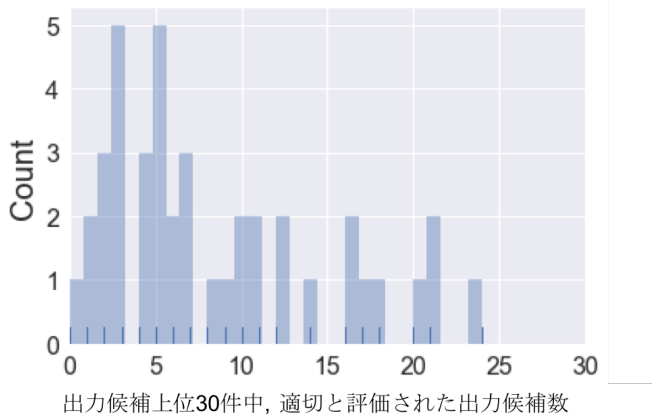


図5 提案手法で失敗した41件について出力候補上位30件までを再評価し、適切と評価されたデータのヒストグラム

や情報提供など発話行為ごとに抽出することが考えられる。特に発話行為ごと抽出について、バックチャネルと質問文を組み合わせた連続応答は聞き役対話システムとして重要である。

文 献

- [1] Higashinaka Ryuichiro, Kobayashi Nozomi, Hirano Toru, Miyazaki Chiaki, Meguro Toyomi, Makino Toshiro, and Yoshihiro Matsuo. Syntactic filtering and content-based retrieval of twitter sentences for the generation of system utterances in dialogue systems. *Proc. IWSDS*, pp. 113–123, 2014.
- [2] 堀口純子. コミュニケーションにおける聞き手の言語行動. 日本語教育, No. 64, pp. 13–26, 1988.
- [3] 今石幸子. 聞き手の行動: あいづちの規定条件. 阪大日本語研究, Vol. 5, pp. 95–109, 1993.
- [4] 福田拓也, 若林啓. 雑談システムにおける twitter データからの統計的バックチャネル応答抽出手法. 人工知能学会論文誌, Vol. 33, No. 1, pp. DSH-H.1–10, 2018.
- [5] 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.
- [6] Kobayashi Yuka, Yamamoto Daisuke, Koga Toshiyuki, Yokoyama Sachie, and Doi Miwako. Design targeting voice interface robot capable of active listening. In *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction*, pp. 161–162. IEEE Press, 2010.
- [7] 山口貴史, 井上昂治, 吉野幸一郎, 高梨克也, Nigel G. Ward, 河原達也. 傾聴対話システムのための言語情報と韻律情報に基づく多様な形態の相槌の生成. 人工知能学会論文誌, Vol. 31, No. 4, pp. C-G31.1–10, 2016.
- [8] Inaba Michimasa and Takahashi Kenichi. Backchanneling via twitter data for conversational dialogue systems. In *The 18th International Conference on Speech and Computer (SPECOM-2016)*, pp. 148–155. Springer, 2016.
- [9] Divesh Lala, Pierrick Milhorat, Koji Inoue, Masanari Ishida, Katsuya Takanashi, and Tatsuya Kawahara. Attentive listening system

with backchanneling, response generation and flexible turn-taking. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue, Saarbrücken, Germany, August 15-17, 2017*, pp. 127–136, 2017.

- [10] M. Schuster and K.K. Paliwal. Bidirectional recurrent neural networks. *Trans. Sig. Proc.*, Vol. 45, pp. 2673–2681, 1997.
- [11] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, Vol. 9, pp. 1735–1780, 1997.