雑談システムにおけるバックチャネル応答の抽出に関する研究

筑波大学
図書館情報メディア研究科
2018年3月
福田　拓也

目 次

第1章　はじめに

近年，音声認識システムの高度化が注目を集めている．音声認識システムはいまやスマートフォンに搭載されているだけでなく，リビングに置かれる家庭用音声アシスタントや，店頭で PR に用いられるロボットなどにも用いられており，これまでのように意識的にシステムを利用する心構えを前提としたシステムではなく，普段の生活の中でユビキタスに利用されるシステムへと活用の状況がシフトしてきている．このような状況の中，音声認識システムは必然的に普段の状況から発せられる雑談に対応する機能が求められており，近年では雑談システムに関する研究が活発に行われている．ここで，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り，人型ロボットへの信頼感を高める効果や，介護や医療において老齢者などとの話し相手としての活用，娯楽目的などの利用が期待されている．

本研究では，特に雑談システムの「聞き役」としての機能に着目する．雑談システムの応答内容について，図 1.1 の「身長をもっと伸ばしたい」というユーザ発話に対する応答を考える．応答例の 1 つとして「伸ばすには成長ホルモンが特に重要になんだって」というような情報提供を行う生産的な応答が挙げられる．しかし，人間同士の会話ではこのような応答ばかりでやりとりされるのではなく,「あ～分かる！」というような一見，非生産的な応答も多く用いられる．

このような，聞き手側が行う会話中の短いリアクションはバックチャネルと呼ばれている．バックチャネルには，相手の話を聞いたり理解していることを伝える役割があることから，コミュニケーションをより円滑なものにする働きがあると言われている [1]．もし，聞き手が一切反応せずに相手の話を黙って聞いていると話し手はたちまち不安になって会話をやめてしまう [2]．また，バックチャネルが常に「うん」や「はい」といった典型的な

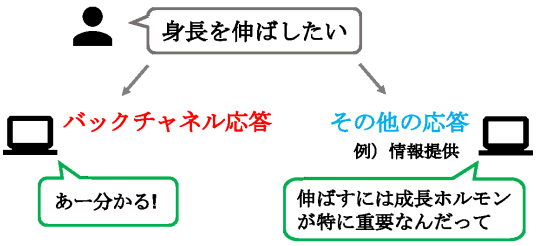

図 1.1: ユーザ発話に対するバックチャネル応答とその他の応答

相槌ばかりの場合でも，話し手に対して話を聞いていない印象を与えてしまい，同様に会話の意欲を失わせてしまうことが考えられる．このため，雑談においてバックチャネルが果たす役割は大きく，多様なバックチャネルを相手の発話に合わせて柔軟に返答することは人間との自然な対話を実現するために重要な要素である．

バックチャネルという用語は，日本語と英語で文化的な意味合いの違いが大きいが，元々は国外で生まれた用語である．Yngve[3] は，バックチャネルとは，発話権が話し手側にある際の聞き手側の「yes」や「uh-huh」等の短いメッセージを意味する用語としている．その後，Duncan ら [4] では短いメッセージの他に，相手の発話に付け加えて文として完成させる表現や内容について問いただす表現，非言語的振る舞い（うなずき等）を作ることが可能な表現として，その解釈を拡大させている．一方，日本語においては，英語よりも頻繁に「うん」,「ええ」,「なるほど」などの “相槌” を用いることから，それとの関係が議論の対象になる．堀口 [5] は，バックチャネルを相槌と対応する概念とした上で，その中にも上記のような “いわゆる相槌” という典型的な表現形式のパターンが存在しており，このような表現形式を “相槌詞” と呼ぶことを提案している．しかし，この提案に対して今石 [6] は，“相槌” の中に “相槌詞” が含まれるという用語の混乱を指摘し，話し手側の発話に対する聞き手側のフィードバック情報全体を “バックチャネル” という用語で表し，“相槌” をバックチャネルに内包される概念として定義している．

本研究では，この今石 [6] 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より広い意味合いでの聞き手側のフィードバックを表す用語として用いる．ある発話がバックチャネルかそうでないかを厳密に線引きすることは難しいが，この定義における基本的な判断基準は，話し手の発話権を奪わないという点にある．この観点からは，例えば「そうなの？」「確かにね」「大丈夫か」といった発話は，バックチャネルと考えられる．

本研究では，以上の観点に基づいて，雑談システムにおいて聞き役として適切なバックチャネルを生成する手法を提案する．これまでに，聞き役として適切な応答を生成する手法はいくつか提案されている [7, 8, 9] が，従来の手法はバックチャネルの種類をあらかじめ固定して決めた上で，分類手法によってその中から適切なものを選択するアプローチを取っている．このため，従来手法で扱うバックチャネルは，比較的典型的な “相槌” に限定されているのが現状である．

一方で，情報提供などを積極的に行うことを目指す雑談システムでは，応答の生成方法として，あらかじめ用意した大規模言語コーパスを応答候補文の集合として，システムの応答に適切な文章を選択する手法が提案されている．応答候補文の集合は 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．本研究では，このアプローチを聞き役としてのバックチャネルを生成する目的に援用し，Twitter データから適切なバックチャネルを抽出する手法を提案する．これにより，ユーザ発話に合わせ

た多様なバックチャネル応答を抽出することを目指す.

第2章　関連研究

バックチャネル応答に関連する研究として，聞き役になることを目的とした “傾聴システム” の研究が行われている．傾聴システムの目的は，ユーザの話したいという欲求を満足させることにある [10]．この目的を達成するため，傾聴システムではバックチャネル応答の生成も行われる [7, 11, 8] が，ユーザの話を促すための関連話題の提供 [12] や，ユーザの話題を掘り下げる質問 [13]，自己開示発話など，様々な対話行為も組み合わせて用いることが考えられる．このアプローチでは，継続的な対話の中で適切な応答タイプを選択する対話制御が重要な観点であり，実際の発話生成では，選択された応答タイプに対応した定型パターンを用いるアプローチが取られることが多い．このため，生成されるバックチャネルは定型的なパターンに従ったものに制限され，会話自体が単調になりやすいと考えられる．これらの研究に対する本研究の貢献は，傾聴システムの中でバックチャネル生成が必要と判断された時に利用できるモジュールとして，多様なバックチャネル生成を実現する手法を提案することにある．

より細かいニュアンスの違いを考慮したバックチャネル生成に着目した研究として，山口ら [9] や Inaba ら [14] が挙げられる．山口ら [9] は応答系感動詞（促しや受容を表す「うん」や「ふんふん」など）や感情表出系感動詞（興味や関心・共感を表す「あー」や「はー」など）を対象とし，さらに応答系についてはその強さによって 3 つのクラスに分けて，適切な応答クラスの予測を行っている．また，Inaba ら [14] は,「大丈夫ですか」や「残念です」,「かっこいいね」などの細分化された 44 種類のバックチャネルの応答タイプを定義し，入力文に対してどの応答クラスが適切であるかという多値分類問題として定式化している．しかし，バックチャネルには「羨ましい！」など相手の発話内容についてある程度踏み込んだ限定的な使われ方をするものもあり，実際の会話で使用されるバックチャネル応答はニュアンスの差も含め多岐にわたる．また，応答タイプの細分化を進めるほど，個々の応答タイプの役割は直交しない場合が多くなり，分類手法を適用するのに必要な教師データの作成は難しくなる．これらの課題を克服するため，本研究では分類手法に帰着させる従来のアプローチではなく，Twitter データからの抽出に基づく手法を提案する．

このほかに関連する研究として，音声での対話を前提とした雑談システムにおけるバックチャネル応答に関する研究では，適切な応答のタイミングについて着目した手法 [15, 16, 17] が提案されている．本研究では，いつバックチャネル応答をするべきかのタイミングについては所与のものとして扱わないが，将来的にはこのような観点も考慮してタイミングを

推定する必要があると考えられる．また，下岡ら [18] は，音声認識誤りを考慮して生成する応答タイプを選択する手法を提案しているが，本稿ではテキスト上の対話を対象として，このような誤りの影響は考えないものとする．

第3章　提案手法

本研究では，入力としてユーザ発話文 u が与えられたときに，応答候補文の集合 S の中から適切なバックチャネルを抽出する手法を提案する．出力がユーザ発話文 u のみに依存して決まるという意味で，提案手法はいわゆる「一問一答」形式での応答抽出手法である．

本研究では条件の異なる 2 つの手法を提案する．提案手法の概要をそれぞれ図 3.1，図 3.2 に示す．本研究では，応答候補文の集合 S は 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．Twitter では，ユーザは Reply 機能によって特定の Tweet に対する返信を行うことができる．提案手法は，この Reply の関係を利用して，バックチャネルとしての適切さを考慮する特徴量を計算する．

後述する特徴量抽出の条件を満たすため，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．つまり，T をコーパスとして利用可能な Tweet 全体の集合，$R(s,t)$ を Tweet t が Tweet s に対する Reply であることを表す述語とすると，応答候補文の集合 S は以下のように定義される．これは Nio ら [19] と同様のコーパス設計である．

$$S = \{t \in T | \exists s \in T[R(s,t)], \exists r \in T[R(t,r)]\} \tag{3.1}$$

Twitter では，一つの Tweet に対して複数の Reply が存在できるが，一つの Tweet が複数の Tweet に対する Reply となることはできない（ $\forall t, s, r[R(s,t), R(r,t) \implies s = r]$ ）．

3.1　応答候補文の特徴量

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 つを提案する．

1. ユーザ発話文 u と応答候補文 t の結束度 $f_1(t,u)$
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 $f_3(t)$

これらの特徴量抽出関数では，単語の IDF (Inverse Document Frequency) を用いる．語彙 w の IDF は，応答候補文集合 S を文書集合，$DF(w)$ を S の中で語彙 w を含む Tweet の数として，あらかじめ以下のように求める．

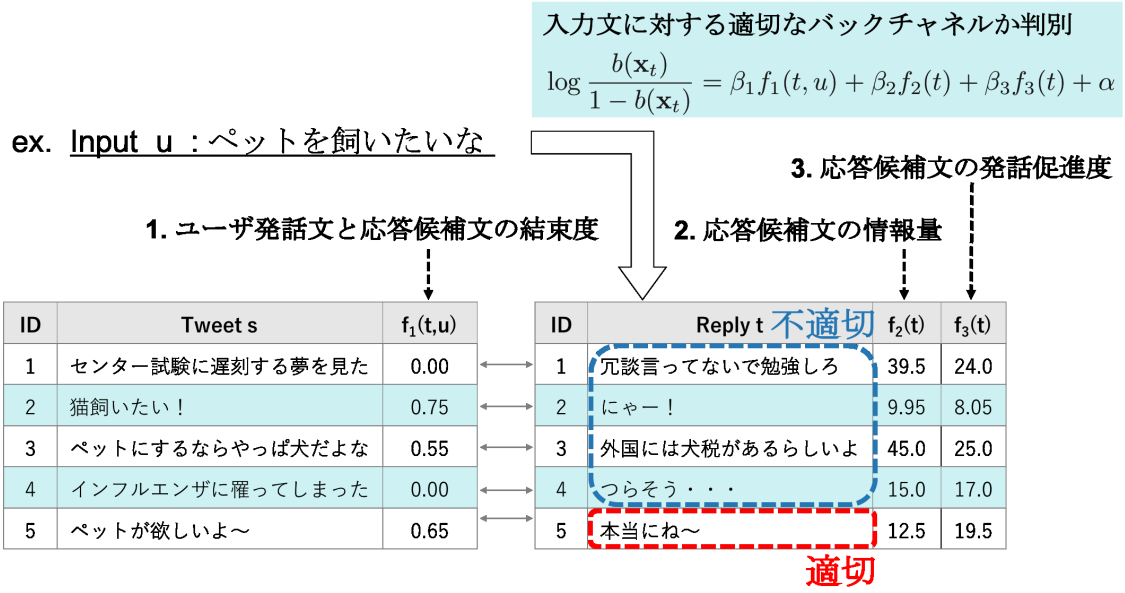

図 3.1: 提案手法（f_1）の概要

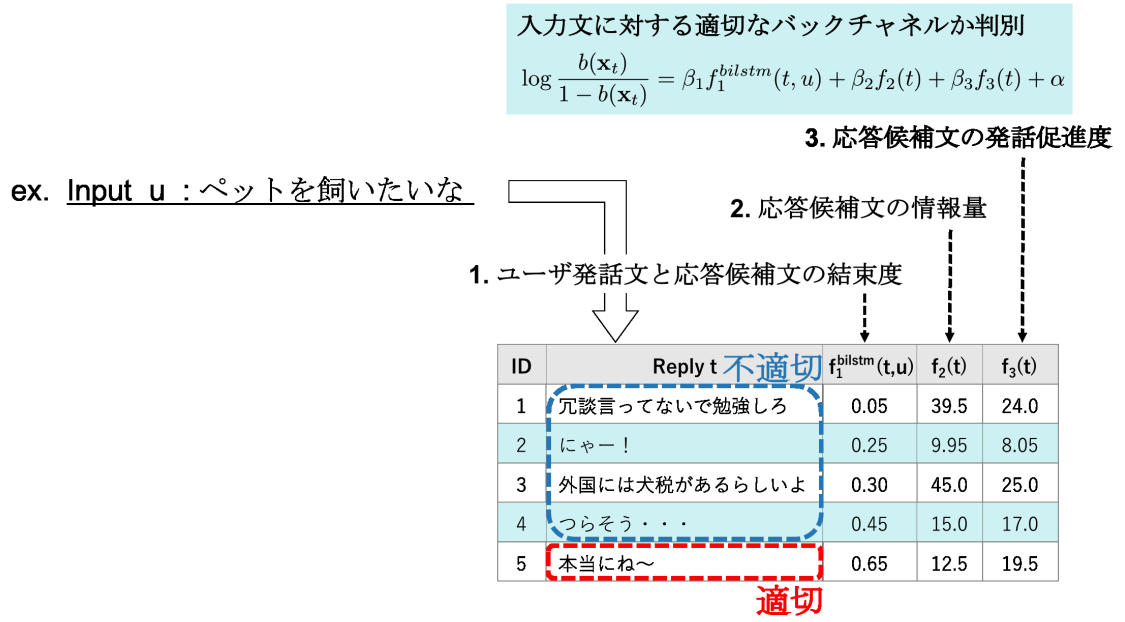

図 3.2: 提案手法（f_1^{bilstm}）の概要

$$IDF(w) = \log \frac{|S|}{DF(w)} \tag{3.2}$$

3.1.1 ユーザ発話文と応答候補文の結束度

Tweet と Reply 関係を用いた結束度推定

応答候補文である Tweet t は，別の（ただ一つの）Tweet に対する Reply である．この Reply 先である Tweet を $g(t) = s$ と表す．ユーザ発話文 u と Tweet s の TF・IDF ベクトルを作成し，それぞれ $\mathbf{v}_u$，$\mathbf{v}_s$ とする．ユーザ発話文 u と応答候補文 t の結束度の強さ $f_1(t, u)$ は，$\mathbf{v}_u$ と $\mathbf{v}_s$ のコサイン類似度と定義する．

$$f_1(t, u) = \frac{\mathbf{v}_u \cdot \mathbf{v}_{g(t)}}{|\mathbf{v}_u||\mathbf{v}_{g(t)}|} \tag{3.3}$$

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と考えられる [20].

BiLSTM を用いた結束度推定

上述の推定方法の他に，近年，自然言語処理において文書分類や機械翻訳 [21]，分散表現獲得 [22] といった様々なタスクで高い性能が報告されている深層学習を適用した結束度推定も行い，上述の推定方法との性能を比較する．

自然言語処理では 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）[23] 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（$\mathbf{h_t} = \overrightarrow{\mathbf{h_t}}||\overleftarrow{\mathbf{h_t}}$）．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）[24] 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 3.4 から式 3.9 で定義される．このとき σ はシグモイド関数，$\mathbf{U}_\star \in R^{H \times H}$，$\mathbf{W}_\star \in R^{H \times E}$ 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），$\mathbf{b}_\star \in R^{H \times 1}$ はバイアス項である．LSTM のアーキテクチャではメモリセル $\mathbf{c}$ および 3 つのゲート（入力：i，忘却 f，出力：o）がある．メモリセルでは

表 3.1: 結束度が高いと判定された例

Input u	Reply t	$f_1^{bilstm}(t,u)$
弟に攻殻勧めたら No thank you された	俺が思うに絶対後で弟後悔するよ	0.79
舌先に出来た口内炎が痛い!ストレスだ	今，回復させるね！ヒール！	0.62
腹減った	おれも	0.74

長期の履歴情報を保持し，履歴を遡っても誤差を伝搬させることができる．入力ゲートではメモリセルの状態を入力ベクトル $\mathbf{x}_t$ によって変化させることができる．忘却ゲートではメモリセルが必要なタイミングで以前の状態を覚えたり忘れたりすることができる．これらのゲートを用いてメモリセルの状態を長期，短期の情報のバランスを調整して更新する．最後に，出力ゲートではメモリセルによる隠れ層への出力の影響を制御することができる．

$$\mathbf{i}_t = \sigma(\mathbf{W_i x}_t + \mathbf{U_i h}_{t-1} + \mathbf{b_i}) \tag{3.4}$$

$$\mathbf{f}_t = \sigma(\mathbf{W_f x}_t + \mathbf{U_f h}_{t-1} + \mathbf{b_f}) \tag{3.5}$$

$$\mathbf{o}_t = \sigma(\mathbf{W_o x}_t + \mathbf{U_o h}_{t-1} + \mathbf{b_o}) \tag{3.6}$$

$$\tilde{\mathbf{c}}_t = \tanh(\mathbf{W_c x}_t + \mathbf{U_c h}_{t-1} + \mathbf{b_c}) \tag{3.7}$$

$$\mathbf{c}_t = \mathbf{i}_t * \tilde{\mathbf{c}}_t + \mathbf{f}_t * \mathbf{c}_{t-1} \tag{3.8}$$

$$\mathbf{h}_t = \mathbf{o}_t * \tanh(\mathbf{c}_t) \tag{3.9}$$

上述の Tweet と Reply 関係を用いた結束度の推定方法は間接的にユーザ発話文と発話候補文との結束度を見ている．これに対し，ユーザ発話文について応答候補文が妥当かを直接見るよう考慮し，Siamese-BiLSTM モデル（図 3.3）を用いてユーザ発話文 u と応答候補文 t の結束度の強さ $f_1^{bilstm}(t,u)$ を求める．このような設計のモデルは質問応答タスクにも用いられている [25]．モデルは Tweet と Reply 用の 2 つの BiLSTM を用いてそれぞれをエンコードし，最終的に出力されるベクトルに対して似ているかどうかを判別する．ユーザ発話文 u と応答候補文 t の結束度の強さ $f_1^{bilstm}(t,u)$ は以下のように定義する．表 3.1 に結束度が高いと判定された例，表 3.2 に結束度が低いと判定された例を示す．

$$f_1^{bilstm}(t,u) = \sigma(\frac{\mathbf{h}_u \cdot \mathbf{h}_t}{|\mathbf{h}_u||\mathbf{h}_t|}) \tag{3.10}$$

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．発話の情報量が多いほど，話し手はその内容を考慮した上

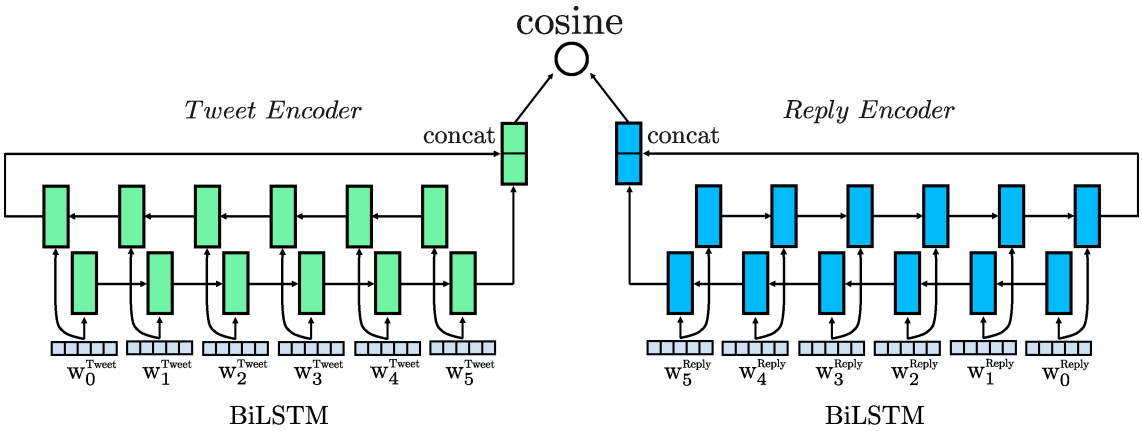

図 3.3: BiLSTM を用いた結束度推定モデル

表 3.2: 結束度が低いと判定された例

Input u	Reply t	$f_1^{bilstm}(t,u)$
俺，女なんだ	早速かい w	0.31
Mova19 年間ありがとう	かぶった w	0.42
面白いアニメがみたいな	おっす!	0.09

で次の発話をしなければならず，発話権を維持することが難しくなる．

ここでは，発話の情報量は，応答候補文 t を形態素解析した結果得られる単語列 $w_{t1}, \cdots, w_{tN}$ に含まれる各単語の情報量の和とする．単語 w の情報量は，情報理論で一般的に用いられる定義に従うと，w を含む Tweet が出現する確率 $p(w)$ に基づいて，$-\log p(w)$ と定義できる．ここで，$p(w)$ の推定に応答候補文の集合 S を用いることにすると，$p(w)$ の最尤推定量は以下のように求められる．

$$p(w) = \frac{DF(w)}{|S|} \tag{3.11}$$

これを情報量の定義に代入すると，以下のように IDF と一致することが分かる [26].

$$-\log p(w) = -\log \frac{DF(w)}{|S|} = \log \frac{|S|}{DF(w)} = IDF(w) \tag{3.12}$$

このため，応答候補文 t の情報量 $f_2(t)$ は，t に含まれる単語の IDF の合計値として求めることができる．

$$f_2(t) = \sum_{i=1}^{N} IDF(w_{ti}) \tag{3.13}$$

これを素性として用いることで，情報量が大き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にできると考えら

れる．

3.1.3 応答候補文の発話促進度

本研究では，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．バックチャネル発話に求められる役割は，話を理解していることを伝え，相手の話を促すことである．この観点からは，図 3.1 の Reply の「にゃー」のように，相手が話を継続しにくいような発話はバックチャネルとして適切ではない．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になっていると考えられることから，本研究では応答候補文の発話促進度として応答候補文の次の発話の平均文字数を特徴量として用いる．

発話候補文 $t \in S$ の発話促進度は，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．Tweet t の文字列を得る関数を $str(t)$ と記述する．また，ある文字列 λ を持つ応答候補文の集合を $Q_\lambda = \{t \in S | str(t) = \lambda\}$ と表すと，発話候補文 t と同一の文字列を持つ応答候補文の集合は $Q_{str(t)}$ と表せる．$Q_{str(t)}$ のいずれかの要素に対する Reply であるような Tweet の集合は，以下のように表される．

$$R_t = \{r \in T | \exists t' \in Q_{str(t)}[R(t', r)]\}$$

文字列 λ から，アルファベット，ひらがな，カタカナ，漢字，句点，句読点以外の文字およびネットスラング「w」が 4 回以上続く文字列を除いた文字数を L_λ とする．応答候補文 t の発話促進度 $f_3(t)$ は，以下のように定義する．

$$f_3(t) = \frac{1}{|R_t|} \sum_{r \in R_t} L_{str(r)} \tag{3.14}$$

表 3.3 は，応答の発話促進度 $f_3(t)$ を実際に計算した結果の例である．この特徴量を用いることで，発話促進度が大きい（平均文字数が大きい）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．

$$\mathbf{x}_t = \begin{pmatrix} f_1(t, u) & f_2(t) & f_3(t) \end{pmatrix}^T \tag{3.15}$$

また，f_1^{bilstm} を用いる場合は以下のようになる．

表 3.3: 応答候補文の発話促進度の例

Tweet t	Reply r の例	$L_{str(r)}$	$f_3(t)$
ウホッ	ウホッウホッ すごくかっこいいですよね	6 12	7.07
んちゅー	はいはいおやすみ ちょ，キスすんなやっ	8 10	8.36
なるほど	それが，めちゃめちゃ苦痛なんどよね 育てた選手が守備職人だったけど評価下がった	17 21	19.3

$$\mathbf{x}_t = \begin{pmatrix} f_1^{bilstm}(t,u) & f_2(t) & f_3(t) \end{pmatrix}^T \tag{3.16}$$

この特徴ベクトルを用いて，ロジスティック回帰によりバックチャネルとして適切か否かを二値分類する．ユーザ発話文 u に対する，応答候補文 t のバックチャネルとしての適切さを表すスコアとして，以下を定義する．

$$b(\mathbf{x}_t) = P(y = 1|\mathbf{x}_t) \tag{3.17}$$

これは素性 $\mathbf{x}_t$ が与えられた場合に，応答文 t が適切なバックチャネルである確率を表す．提案手法では定義した $b(\mathbf{x}_t)$ の推定にロジスティック回帰を用いる．ロジスティック回帰は 2 クラス分類問題における一般化線形モデルの一種であり，式（3.18）で特徴ベクトルをモデル化する．

$$\mathrm{logit}\{b(\mathbf{x}_t)\} = \log \frac{b(\mathbf{x}_t)}{1 - b(\mathbf{x}_t)} = \alpha + \beta^T \mathbf{x}_t \tag{3.18}$$

また，ロジスティック回帰の確率モデルを以下に示す．

$$b(\mathbf{x}_t) = \frac{\exp(\alpha + \beta^T \mathbf{x}_t)}{1 + \exp(\alpha + \beta^T \mathbf{x}_t)} \tag{3.19}$$

この確率が最大であるような応答候補文 t を，提案手法の出力とする[1]．

[1]式（3.19）の最大化と式（3.18）の最大化は等価である．

第4章　評価実験

実験では，既存のアプローチに対する提案手法の有効性を検証する．

4.1　実験用データ

4.1.1　応答候補文の集合と前処理

今回の実験では，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うため Tweet または Reply にハッシュタグや URL を含むデータおよび，あらかじめ定型的な挨拶はバックチャネル応答ではないとし，「おはよう」「おやすみ」「おかえり」を含む Reply をもつデータは除いた．また，Tweet と Reply のどちらかが「。」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．以上によって収集した，Tweet が紐づけられた約 1,500 万件の Reply データを応答候補文の集合 S として用いた．

4.1.2　BiLSTM を用いた結束度推定の学習データ

学習データとして上述の Tweet 収集期間には含まれない期間の 350 万件の Tweet と Reply からなるペアデータを使用した．このペアデータについて，元のペアを正例とし，Tweet に対してランダムに選択した Reply からなるペアを負例とした．なお，ランダムに選択した Reply が元の Reply と同一だった場合，再び選択し直すものとした．実験では 1 件の正例に対して 2 件の負例を作成し，教師ラベルとして正例を 1，負例を 0 とした．

4.1.3　モデルの学習用データ

バックチャネル抽出に用いるロジスティック回帰モデルの学習用データを作成するため，上述の Tweet 収集期間には含まれない期間の Tweet データを用意する．この中から，人手によって，話し手らしい発話であり，かつなるべく互いに異なる話題を含んだ文章になるように 100 件の Tweet を選択し，これを学習データ作成用の入力発話の集合 U' とする（$|U'| = 100$）．それぞれの入力発話 $u \in U'$ に対して，応答候補文の集合 S の中から，特徴量 $f_1(t, u)$ が最も大きい上位 10,000 件の応答候補文 $t \in S$ を抽出し，u に対する応答候

表 4.1: 各特徴量の標準化偏回帰係数（f_1）

variable	coef	std err	z	P > $\|z\|$
const	-2.180	0.174	-12.528	0.000
$f_1(t,u)$	0.790	0.105	7.559	0.000
$f_2(t)$	-0.900	0.253	-3.557	0.000
$f_3(t)$	0.209	0.123	1.696	0.090

補文の集合 $S'_u \subset S$ とする（$|S'_u| = 10,000$）．ここから，3 章で述べた各特徴量に対応する以下の基準により，それぞれの $u \in U'$ に対して 6 つの応答候補文を選択する．

1. S'_u の中で，$f_1(t,u)$ が最も大きい Tweet と，最も小さい Tweet
2. S'_u の中で，$f_2(t)$ が最も大きい Tweet と，最も小さい Tweet
3. S'_u の中で，$f_3(t)$ が最も大きい Tweet と，最も小さい Tweet

また，f_1 の代わりに f_1^{bilstm} を使用する際は，今回，上述の基準 1 を S'_u の中で，$f_1^{bilstm}(t,u)$ が最も大きい Tweet と，最も小さい Tweet として応答候補文を選択した．以上の処理により，入力発話と応答文のペアが 600 件得られる．この 600 件のペアそれぞれについて，第 1 著者が入力発話と応答文を見て，応答文が入力発話に対して適切なバックチャネルかどうかを判断し，適切ならば正例として 1 を，不適切ならば負例として 0 を教師ラベルとした．このアノテーションは，以下の条件を両方満たしているかどうかを基準に判断して行った．

- 入力文とかみ合った応答か
- 発話権が移動せず，入力側が前の発話から続けて応答ができるか

これを教師データとして，LIBLINEAR[27] に実装された座標降下法を用いてロジスティック回帰のパラメータを最適化した．

結束度推定に f_1 を用いる場合の各特徴量についての標準化偏回帰係数を図 4.1 に示す．このとき const（定数項）は-2.180，$f_1(t,u)$ は 0.790，$f_2(t)$ は-0.900，$f_3(t)$ は 0.209 であった．これを見ると $f_2(t)$ が最も影響が大きく，その次に $f_1(t,u)$，$f_3(t)$ となった．そして，f_1^{bilstm} を用いる場合の標準化偏回帰係数を図 4.2 に示す．このとき const は-2.180，$f_1^{bilstm}(t,u)$ は 1.349，$f_2(t)$ は-0.568，$f_3(t)$ は 0.134 であった．これを見ると $f_1^{bilstm}(t,u)$ が最も影響が大きく，その次に $f_2(t)$，$f_3(t)$ となった．また，それぞれ $f_1(t,u)$ と $f_1^{bilstm}(t,u)$ と $f_3(t)$ は値が大きいほど，$f_2(t)$ は値が小さいほどスコアが高くなることが判断できる．

4.1.4 評価用データ

評価用データを作成するために，学習用データと同様に，応答候補文の集合 S に含まれない期間の Tweet データから，話し手らしい発話であり，かつなるべく互いに異なる話題を含

表 4.2: 各特徴量の標準化偏回帰係数（f_1^{bilstm}）

variable	coef	std err	z	P > $\|z\|$
const	-2.220	0.176	-12.599	0.000
$f_1^{bilstm}(t,u)$	1.349	0.134	10.097	0.000
$f_2(t)$	-0.568	0.182	-3.113	0.002
$f_3(t)$	0.134	0.121	1.106	0.269

んだ文章になるように 1,000 件の Tweet を選択し，入力発話の集合 U とした（$|U| = 1,000$）．このとき，入力発話として不快な内容の文章は除いた．

4.1.5 比較手法

評価用データに含まれるそれぞれの入力発話 $u \in U$ に対して，以下の 4 つの手法によって応答文を生成する．

提案手法（f_1） 応答候補文の集合 S の中から，特徴量 $f_1(t,u)$ が最も大きい 10,000 件の応答候補文 $t \in S$ を抽出し，u に対する応答候補文の集合 $S_u \subset S$ とする．提案手法（f_1）は，S_u に含まれる Tweet 全てについて特徴量ベクトル $\mathbf{x}_t$ を求めた上で，バックチャネルスコア式（3.19）が最も大きい候補文を出力とする．

提案手法（f_1^{bilstm}） 提案手法（f_1）の場合と同様の応答候補文の集合 S_u を用いて，これに含まれる Tweet 全てについて特徴量ベクトル $\mathbf{x}_t$ を求める．f_1 の代わりに f_1^{bilstm} を用いてスコアを計算し，最も大きい候補文を出力とする．

f_1 のみ + 上限 10 文字 本研究の主たる仮説は，特徴量 $f_2(t), f_3(t)$ によって，応答のバックチャネルらしさを考慮できることにある．このため，ベースラインとして，10 文字以下の応答はバックチャネルらしいと判断する単純な比較手法を考える．この手法では，記号などを除いた文字数 L_t が 10 以下であるという条件を満たす応答候補文のうち，特徴量 $f_1(t,u)$ が最も大きい応答候補文 $t \in S$ を出力する．

LSTM + 上限 10 文字 多クラス分類に基づく既存のバックチャネル応答手法として，[14] を元にした手法を比較手法として用いる．これは入力発話を LSTM（Long Short Term Memory）を用いてエンコードし，設定したバックチャネルクラスに分類するものである．ただし，本研究では多様なバックチャネル応答の生成を目的とするため，記号などを除いた文字数が 10 以下で，かつ出現頻度が 100 以上の発話をコーパスから全て抽出し，分類クラスの集合とした．形式的には，文字列 λ が分類クラスである条件は以下のように表せる．

- $L_\lambda \leq 10$

- $|Q_\lambda| \geq 100$

これらの条件を満たす発話文字列を全て収集した結果，5,838 のクラスが得られた．モデルに与えるラベル付き学習データには，それぞれの応答候補文の Reply 先の Tweet を用いることができる．すなわち，クラス λ のラベルを持つ学習データの集合は，$\{str(s)|\exists t \in S[R(s,t), t \in Q_\lambda]\}$ で与えられる．

[14] ではこのモデルを 44 クラスの分類に適用しているが，本手法では 5,838 クラスの分類に適用している．各クラス最低 100 件の学習データでは適切な学習が行えていない可能性が懸念されるが，10 分割交差検証を行った結果，正解率は 9.141 %であり，全ての入力発話に対して学習データ中で最も頻度の高いクラスを割り当てる majority-baseline の 0.123 %よりも良いことを確認した．

4.2 評価方法

4.2.1 入力発話に対するバックチャネル応答としての適切さ

実験では，評価用の入力発話に対する応答を「○：適切」,「△：適切とも不適切とも言えない」,「×：不適切」の三段階で被験者に評価してもらう．雑談システムの評価は，被験者の主観による部分が大きく，適切と判断する基準をコントロールすることが難しいと言われていることから，手法の評価値を求める際には以下の手順によって一致度の大きい結果のみを利用することとした．

- 被験者全員が同じ評価をつけたデータ以外を評価時のノイズを考慮するため除去
- ○を 2 点，△を 1 点，×を 0 点として各データのスコアに変換
- 全てのデータのスコアの平均値を，手法の評価値として使用

実験はクラウドソーシングサービスの Lancers[1] を利用し，1 つのデータ（評価データと出力したシステム応答のペア）に対して 3 名に評価をしてもらう．これにより手法ごとに計 3,000 件の実験データを用いて評価する．

なお，被験者には次のような説明文と共に入力文に対する複数の応答例を示し，実験に参加してもらった．

[1] http://www.lancers.jp

今回やって頂くことは「適切なバックチャネルの評価」です.
バックチャネルとは,「大変だね」や「確かにね」「大丈夫ですか？」「かっこいい！」など“聞き手が行う短いリアクション”のことです.
例えば「身長をもっと伸ばしたい」に対して「身長を伸ばすには成長ホルモンが重要だそうです」という応答はバックチャネルとしては不適切で,「あ～分かる！」のような聞き役らしい応答をバックチャネルとして適切と判断してください.
入力文と応答文のペアがいくつか表示されますので，それぞれの応答が適切なバックチャネルかどうかを「○，△，×」で評価してください.

例）
-=-
入力文：日本のアニメのクオリティは高いよね！

応答例１：確かにね
応答例１の評価：○（正しい）

応答例２：特に私はジブリ作品が好きです
応答例２の評価：×（会話は噛み合っているが，バックチャネルではない）

応答例３：よろしく
応答例３の評価：×（バックチャネルだが，会話が噛み合っていない）

応答例４：ん～
応答例４の評価：△（判断に困った場合）
-=-

4.2.2　応答文単体でのバックチャネルらしさ

各手法が生成した全ての応答について，応答文単体で見たときにバックチャネルらしい発話であるかどうかを，クラウドソーシングにより評価する．被験者には，いくつかバックチャネルの例を示した上で，各応答文を単体で見てもらい,「○：バックチャネルらしい」,「△：どちらとも言えない」,「×：バックチャネルらしくない」の三段階で判定してもらう．被験者の判定する応答文にバックチャネルらしくない応答文もある程度の割合で含まれるよう，評価対象の応答文と同じ数だけ応答候補文の集合 S の中からランダムに選択し，被験者に提示した．

「入力発話に対するバックチャネル応答としての適切さ」の評価で行った方法と同様に，1 つの応答文に対して 3 名に評価してもらい，被験者全員が同じ評価をつけた応答文について○を 2 点，△を 1 点，×を 0 点としてスコアに変換する．このスコアに基づいて，各手法が出力した応答文がどれだけバックチャネルらしいかを評価する．

4.2.3 応答の多様さ

各手法が生成した応答の多様さを評価するため，出力された応答の文字列としての異なり数を確認する．ある手法 ζ によって得られた，入力発話と出力の応答文字列のペアの集合を $O_\zeta = \{(u_1, \lambda_1), ..., (u_{1000}, \lambda_{1000})\}$ とする（$|O_\zeta| = |U| = 1,000$）．出力の応答の文字列としての異なり数は $|\{\lambda|(u, \lambda) \in O_\zeta\}|$ と表すことができるが，これを出力した延べ発話数で除算した以下の値を応答の多様さの指標とする．

$$diversity(\zeta) = \frac{|\{\lambda|(u, \lambda) \in O_\zeta\}|}{|O_\zeta|} \tag{4.1}$$

4.2.4 入力発話への踏み込み具合

提案手法の狙いは，ニュアンスの違いを考慮した多様な応答を行うことで，入力発話に対してより踏み込んだ応答を可能にすることにある．ここでは，応答の踏み込み具合を定量的に考察するため，コーパスにおける応答文字列の頻度を確認する．入力発話に対して適切と判定された応答であっても，その応答文字列が頻出であれば，当たり障りがなくあまり踏み込んでいない応答である可能性が高いと考えられる．

手法 ζ による入力発話と出力の応答文字列のペアの集合 O_ζ のうち,「入力発話に対するバックチャネル応答としての適切さ」の評価において適切と判断された（3 名全員が○と判定した）ものの集合を $O'_\zeta \subset O_\zeta$ とする．それぞれの応答文字列 λ の出現頻度 $|Q_\lambda|$ の和を求め，適切と判断された応答の数で除算した以下の値を，当たり障りのなさを表すスコアとする．このスコアが小さいほど，より踏み込んだ応答をして適切と判断されている傾向があるといえる．

$$typicalness(\zeta) = \frac{\sum_{(u,\lambda) \in O'_\zeta} |Q_\lambda|}{|O'_\zeta|} \tag{4.2}$$

4.3 結果と考察

「入力発話に対するバックチャネル応答としての適切さ」について，実験の結果を図 4.1 に示す．提案手法（f_1）では 3 名の評価者全員が同じ評価をしたデータは 461 件あった．その中で全員が「○：適切」としたデータは 273 件（59.2%）,「△：適切とも不適切とも言えない」が 19 件（4.1%）,「×：不適切」が 169 件（36.7%）となった．提案手法（f_1^{bilstm}）では 3 名の評価者全員が同じ評価をしたデータは 461 件あった．その中で全員が「○：適切」としたデータは 294 件（72.8%）,「△：適切とも不適切とも言えない」が 9 件（2.2%）,「×：不適切」が 101 件（25.0%）となった．また，f_1 のみ + 上限 10 文字の比較手法では 3 名の評価者全員が同じ評価をしたデータは 494 件あった．その中で全員が「○：適切」としたデータは 260 件（52.6%）,「△：適切とも不適切とも言えない」が 7 件（1.4%）,「×：

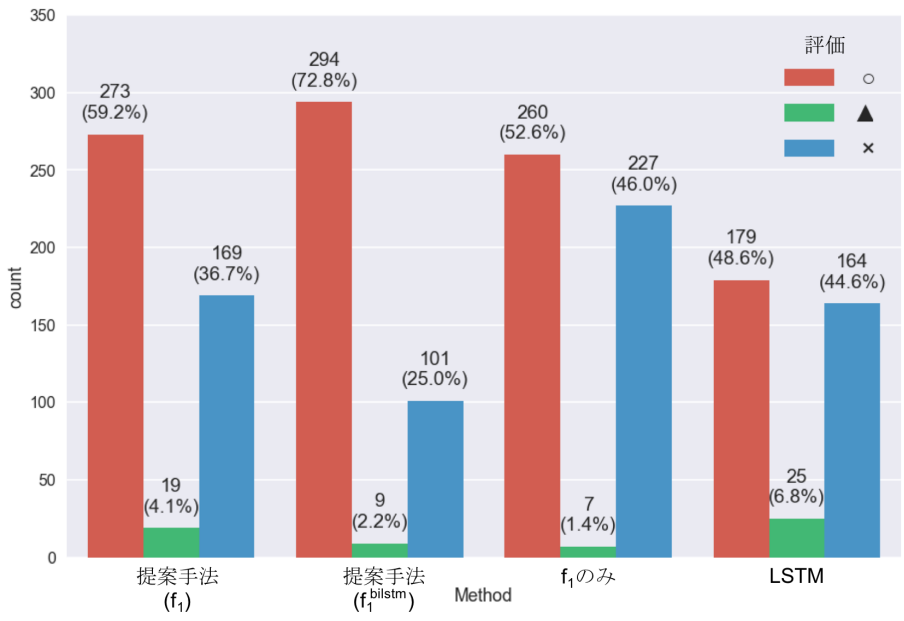

図 4.1: 各手法による入力発話に対するバックチャネル応答としての適切さの評価結果

不適切」が 227 件（46.0%）となった．LSTM + 上限 10 文字の比較手法では 3 名の評価者全員が同じ評価をしたデータは 368 件あった．その中で全員が「○：適切」としたデータは 179 件（48.6%）,「△：適切とも不適切とも言えない」が 25 件（6.80%）,「×：不適切」が 164 件（44.6%）となった．評価の一致率について，Fleiss の κ 係数 [28] を適用すると 0.362 となり，低い一致であった（poor agreement）．

次に，手法ごとの平均評価値を表 4.3 に示す．提案手法（f_1）では平均評価値は 1.226，標準偏差は 0.954 であった．提案手法（f_1^{bilstm}）では平均評価値は 1.478，標準偏差は 0.867 であった．また，f_1 のみの比較手法では平均評価値は 1.067，標準偏差は 0.992 であった．LSTM による比較手法では平均評価値は 1.041，標準偏差は 0.966 であった．2 項検定を行った結果，提案手法（f_1）と f_1 のみの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=0.012）．また，提案手法（f_1）と提案手法（f_1^{bilstm}），提案手法（f_1）と LSTM による比較手法についてもそれぞれ同様に有意であった（p 値 < 0.001，p 値=0.006）．提案手法（f_1^{bilstm}）と f_1 のみの比較手法，提案手法（f_1^{bilstm}）と LSTM による比較手法も有意であった（p 値 < 0.001）．このことから，提案手法では比較手法よりも適切な応答を行うことができているといえる．また，特徴量抽出関数を Tweet と Reply 関係を用いて間接的にユーザ発話文と発話候補文との結束度を見ている f_1 から，ユーザ発話文について応答候補文が妥当かを直接見るよう考慮した f_1^{bilstm} にすることでより適切な応答を期待することができるといえる．

次に「バックチャネルらしさ」について，手法ごとの評価値を表 4.4 に示す．提案手法（f_1）では平均評価値は 1.951，標準偏差は 0.291 であった．提案手法（f_1^{bilstm}）では平均

表 4.3: 入力発話に対するバックチャネル応答としての適切さの平均評価値．提案手法（f_1）と提案手法（f_1^{bilstm}）に有意差あり（p 値 < 0.001）．提案手法（f_1）と f_1 のみの比較手法にも有意差あり（p 値=0.012）．また，提案手法（f_1）と LSTM の比較手法にも有意差あり（p 値=0.006）．提案手法（f_1^{bilstm}）と f_1 のみの比較手法，提案手法（f_1^{bilstm}）と LSTM による比較手法にも有意差あり（p 値 < 0.001）．

手法	平均評価値	標準偏差	件数
提案手法（f_1）	1.226	0.954	461
提案手法（f_1^{bilstm}）	**1.478**	0.867	404
f_1 のみ	1.067	0.992	494
LSTM	1.041	0.966	368

表 4.4: バックチャネルらしさの平均評価値．提案手法（f_1）と提案手法（f_1^{bilstm}）に有意差なし（p 値=0.626）．提案手法（f_1）と f_1 のみの比較手法に有意差あり（p 値 < 0.001）．また，LSTM の比較手法と提案手法（f_1）にも有意差あり（p 値 < 0.001）．提案手法（f_1^{bilstm}）と f_1 のみの比較手法，提案手法（f_1^{bilstm}）と LSTM による比較手法にも有意差あり（p 値 < 0.001）．

手法	平均評価値	標準偏差	件数
提案手法（f_1）	1.951	0.291	586
提案手法（f_1^{bilstm}）	1.959	0.276	488
f_1 のみ	1.834	0.527	416
LSTM	**2.000**	0.000	858

評価値は 1.959，標準偏差は 0.276 であった．また，f_1 のみの比較手法では平均評価値は 1.834，標準偏差は 0.527 であった．LSTM による比較手法では平均評価値は 2.0，標準偏差は 0.0 であった．2 項検定を行った結果，提案手法（f_1）と f_1 のみの比較手法の差は有意であり，LSTM による比較手法と提案手法（f_1）の差についても同様に有意であった（p 値 < 0.001）．提案手法（f_1^{bilstm}）と f_1 のみの比較手法，提案手法（f_1^{bilstm}）と LSTM による比較手法も有意であった（p 値 < 0.001）．提案手法（f_1）と提案手法（f_1^{bilstm}）に有意差はみられなかった（p 値=0.626）．また，応答の多様さと，当たり障りのなさを表す指標のスコアを表 4.5 に示す．diversity のスコアは高いほど応答が多様であり，typicalness のスコアは低いほど入力発話に対する踏み込み具合が高い傾向にあり望ましいといえる．

LSTM による比較手法は，4 つの手法の中で最もバックチャネルらしい応答を生成する傾向が強いが，応答の多様さの度合いと踏み込み具合は非常に弱い．これは，LSTM による比較手法の手法が「えっ」などの頻度が高く幅広い入力発話に対して使える応答を頻繁に行うことに起因している．このことから，分類手法に基づく従来のアプローチを，クラス数を非常に大きくして多様なバックチャネル応答に適用することは難しいという本研究の主張は，支持されていると考えられる．

表 4.5: 応答の多様さと踏み込み具合に関する指標

手法	diversity	typicalness
提案手法（f_1）	0.832	3206.90
提案手法（f_1^{bilstm}）	0.854	2727.00
f_1 のみ	**0.961**	**1455.93**
LSTM	0.036	12735.53

一方，f_1 のみの比較手法は，抽出に基づくアプローチであることから，多様さの度合いと踏み込み具合が非常に強い．しかし，バックチャネルらしさの評価値は，10 文字以下の発話のみを用いることによってある程度高くなっているものの，提案手法との比較では有意に低い．また，入力発話に対するバックチャネル応答としての適切さの評価においても，3 名の被験者から適切でないと判断されるような応答が提案手法と比べて多いことがわかる．この比較から，提案手法は多様さの度合いと踏み込み具合が強いという抽出に基づくアプローチの特長を保ちながら，素性 f_2, f_3 によって効果的にバックチャネルらしい適切な応答を抽出することができていると結論づけることができる．

最後に，提案手法が適切な応答ができた例と，適切でない応答を行った例をそれぞれ表 4.6，表 4.7 に示す．失敗例の多くは，評価データに対してかみ合った応答ができていないケースである．この原因として，発話候補文の発話促進度 $f_3(t)$ を考慮することにより抽出されるバックチャネルに偏りが生まれてしまったことが挙げられる．例えば，提案手法では特に “大丈夫” を含む応答が多く，提案手法（f_1）では 1,000 件の評価データに対して 130 件，提案手法（f_1^{bilstm}）では 99 件であったのに対し，f_1 のみの比較手法では 39 件であった．不適切な応答をしたケースでは，このような入力発話との繋がりとは独立に選ばれやすい応答を抽出している場合が多くみられた．このため，本研究では線形モデルを用いたが，今後は非線形な表現能力を持つモデルの適用を検討する必要があると考えられる．

また，f_1 を用いた場合では，他の原因としては評価データに対してそもそも類似度が高いデータがなかったということもあるが，大きくはハイコンテキストなやり取りがなされているデータを考慮していなかったことが考えられる．ハイコンテキストな応答の例を表 4.8 にいくつか示す．例えば,「俺，女なんだ」という Tweet に対する Reply として「早速かい w」とあり，これは一見するとつじつまの合っていない応答と感じてしまう．実はこのやり取りは “4 月 1 日の午前中” になされたものであり，このことからこの Reply はエイプリルフールでの冗談の Tweet に対してなされた応答であると判断できる．しかし，これは “4 月 1 日の午前中” という日時の情報があって初めて解釈できるわけであり，そのような情報を考慮していない提案手法（f_1）にとってはノイズデータであり，性能は大きく影響されてしまう．このことが原因で，結果的に提案手法（f_1）では応答としてつじつまの合わないものを選んでしまう可能性がある．このため，ハイコンテキストなやり取りがな

表 4.6: 提案手法の成功例

評価データ	手法	応答文	評価の内訳
新幹線に山ピー発見!!!!!!!	提案手法（f_1） 提案手法（f_1^{bilstm}） f_1 のみ LSTM	まじで !? ｗｗ すげー 愛、テキサス おめでとう!	○: 3 ○: 3 × : 3 ○: 1 × : 2
気持ちの切り替えって大事よね	提案手法（f_1） 提案手法（f_1^{bilstm}） f_1 のみ LSTM	大事！ ほんまそれ なんくるないさ〜 なでなで	○: 3 ○: 3 × : 3 △: 1 × : 1
ヘリの音っていいよね	提案手法（f_1） 提案手法（f_1^{bilstm}） f_1 のみ LSTM	た、たしかに ww うるさいよー お邪魔してます えっ	○: 3 ○: 3 × : 3 ○: 2 × : 1
自分が情けなくて泣いてしまいそう。	提案手法（f_1） 提案手法（f_1^{bilstm}） f_1 のみ LSTM	大丈夫か つらそう ぎゅっぎゅ 大丈夫?	○: 3 ○: 3 △: 1 × : 2 ○: 3
俺頭悪すぎクソワロタ wwwwww	提案手法（f_1） 提案手法（f_1^{bilstm}） f_1 のみ LSTM	詳細ｋｗｓｋ ドンマイ なんだ嘘か。 大丈夫?	○: 3 ○: 3 ○: 1 × : 2 ○: 3
週に一回焼き肉食べないとだめ	提案手法（f_1） 提案手法（f_1^{bilstm}） f_1 のみ LSTM	わかる!! 食べたーい 一昨日食べた！ えっ	○: 3 ○: 2 △: 1 ○: 1 △: 1 × : 1 ○: 1 △: 1 × : 1

表 4.7: 提案手法の失敗例

評価データ	手法	応答文	評価の内訳
白湯効果すごい	提案手法（f_1）	それだ	×：3
	提案手法（f_1^{bilstm}）	おちつく〜	△: 1 ×：2
	f_1 のみ	ロック	×：3
	LSTM	おはよう	×：3
ヨーロッパと南米はサッカー界の中心です	提案手法（f_1）	いいなー	×：3
	提案手法（f_1^{bilstm}）	がんばってね	×：3
	f_1 のみ	いいなー	×：3
	LSTM	えっ	△: 2 ×：1
貰った本とか読んだ	提案手法（f_1）	どうした	×：3
	提案手法（f_1^{bilstm}）	面白い！	×：3
	f_1 のみ	だうとだうと	△: 1 ×：2
	LSTM	おめ	×：3
不健康自慢はやめろぉ!	提案手法（f_1）	可愛いw	×：3
	提案手法（f_1^{bilstm}）	お大事に	×：3
	f_1 のみ	ワロス	○: 1 △: 1 ×：1
	LSTM	えっ	△: 2 ×：1
プール行きたい!海行きたい!泳ぎたい!!	提案手法（f_1）	酔ってる www	×：3
	提案手法（f_1^{bilstm}）	泳げる？	△: 2 ×：1
	f_1 のみ	海行きたい	○: 2 ×：1
	LSTM	それな	○: 1 △: 2
英語って最初わかんなくても何回か読めばだんだんわかってくるね。	提案手法（f_1）	大丈夫？	×：3
	提案手法（f_1^{bilstm}）	ムズすぎ www	○: 1 △: 1 ×：1
	f_1 のみ	がんば	○: 2 △: 1
	LSTM	えっ	×：3

表 4.8: ハイコンテキストな応答例

Tweet	Reply
俺，女なんだ	早速かい w
mova19 年間ありがとうお疲れ様	かぶった w
今からゲームでもしようかな	いや さっさと寝ろよ！
面白いアニメが見たいな	おっす!

されているデータかどうかを分類してそのようなデータをできるだけ除く必要がある．これに対して，f_1^{bilstm} を用いた場合では，ユーザ発話と応答候補文を直接評価しているため元のやり取りを考慮する必要がなく，実験の結果から BiLSTM を用いた結束度推定の有用性を示している．

また，今回の実験では顕著な問題としては観測されなかったが，傾聴システムの研究では，適切な応答のためには入力発話の対話行為の考慮が重要であることが指摘されている [10]．この観点を考慮することで，質問の発話に対しては「マジで？」や「確かに」といったバックチャネル応答では対話が成立しないなどの判断がより適切に行える可能性がある．特に，本実験では一問一答の応答での適切さを評価しているが，傾聴システムから提案手法を活用するなど長期の対話の中での応答を行った時に対話行為や文脈などの考慮が一層重要になることも予想されるため，より一般的な状況で提案手法の有効性を評価することは今後の課題といえる．

第5章　おわりに

本研究ではバックチャネルの種類を制限せず，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手法を提案した．提案手法は，ユーザ発話文と応答候補文の結束度，応答候補文の情報量，応答候補文の発話促進度の 3 つの特徴量からなる特徴ベクトルからバックチャネルらしさを表すスコアを求めることで，適切なバックチャネル応答を行う．実験では，入力文に対して適切なバックチャネルを出力できたかどうかを，分類ベースに基づく手法や類似度と文字数制限に基づく単純な手法と比較した．この結果，被験者全員が同じ評価をしたデータについて提案手法の方が有意に高い評価を得ることができ，提案手法の方が不適切である応答をしにくいことを明らかにした．

課題としては，4.3 でも述べたように発話候補文の発話促進度 $f_3(t)$ を考慮することにより抽出されるバックチャネルに偏りが生まれてしまうことがある．また，バックチャネルには相手の話の補完や繰り返し，さらには言い換え表現としてのパターンも存在する．本研究では，基本的に文章としての情報量が少ないものがより適切なバックチャネルらしいとしているため，このような表現に対して提案手法は必ずしもうまく抽出できるわけではない．このため，例えば Tweet と Reply が似た文章かなど，Tweet と Reply がどのような関係にあるかを考慮する必要があると考えられる．

今後の展望として，バックチャネルと情報提供などを組み合わせた応答生成が検討できる．1.1 では，ユーザ発話「身長をもっと伸ばしたい」に対するシステムの応答として,「伸ばすには成長ホルモンが特に重要になんだって」という情報提供を行う応答と「あ～分かる！」というバックチャネル応答を例として挙げたが,「あ～分かる！伸ばすには成長ホルモンが特に重要になんだって」というバックチャネル応答の後に情報提供を行う応答も十分に考えられる．バックチャネル応答と情報提供を行う応答を個別のモジュールで扱うとした際，バックチャネル応答モジュールとして提案手法を用いることができる．このような応用を考える際には，出力されるそれぞれの応答について自然なペアを判定する方法が重要になると考えられる．

参考文献

[1] 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.

[2] 水谷信子. あいづち論. 日本語学, Vol. 7, No. 13, pp. 4–11, 1988.

[3] Victor H. Yngve. On getting a word in edgewise. In *Papers from the 6th Regional Meeting of the Chicago Linguistic Society*, pp. 567–578. Univ. of Chicago Dept. of Linguistics, 1970.

[4] Starkey Duncan and Donald W. Fiske. *Face-to-Face Interaction: Research, Methods, and Theory*. L. Erlbaum Associates, 1977.

[5] 堀口純子. あいづち研究の現段階と課題. 日本語学, Vol. 10, No. 10, pp. 31–41, 1991.

[6] 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.

[7] 井上昂治, 河原達也. 自律型アンドロイド erica のための音声対話システム. 言語・音声理解と対話処理研究会, Vol. 75, pp. 21–24, 2015.

[8] 佐藤康将, 目良和也, 市村匠, 山下利之, 相沢輝昭, 吉田勝美. 肯定/否定意図を検出するチャットシステムのためのあいづち生成手法. ファジィシステムシンポジウム講演論文集, Vol. 17, pp. 509–512, 2001.

[9] 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.

[10] 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.

[11] Yuka Kobayashi, Daisuke Yamamoto, Toshiyuki Koga, Sachie Yokoyama, and Miwako Doi. 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.

[12] 大竹裕也, 萩原将文. 高齢者のための発話意図を考慮した対話システム. 日本感性工学会論文誌, Vol. 11, No. 2, pp. 207–214, 2012.

[13] 石田真也, 井上昂治, 中村静, 高梨克也, 河原達也. 傾聴対話システムのための発話を促す聞き手応答の生成. 言語・音声理解と対話処理研究会, Vol. B5, No. 01, pp. 1–6, 2016.

[14] Michimasa Inaba and Kenichi Takahashi. Backchanneling via twitter data for conversational dialogue systems. In *The 18th International Conference on Speech and Computer (SPECOM-2016)*, pp. 148–155. Springer, 2016.

[15] 西村良太, 北岡教英, 中川聖一. 応答タイミングを考慮した雑談音声対話システム. 言語・音声理解と対話処理研究会, Vol. 46, pp. 21–26, 2006.

[16] 大野誠寛, 神谷優貴, 松原茂樹. タグ付けの安定性を備えた音声対話コーパスに基づくあいづち生成タイミングの検出. 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション, Vol. 110, No. 356, pp. 19–24, 2010.

[17] Nigel Ward and Tsukahara Wataru. Prosodic features which cue backchannel responses in english and japanese. *The Journal of Pragmatics*, Vol. 32, No. 8, pp. 1177–1207, 2000.

[18] 下岡和也, 徳久良子, 吉村貴克, 星野博之, 渡部生聖. 音声対話ロボットのための傾聴システムの開発. 自然言語処理, Vol. 24, No. 1, pp. 3–47, 2017.

[19] Lasguido Nio, Sakriani Sakti, Graham Neubig, Tomoki Toda, and Satoshi Nakamura. Conversation dialog corpora from television and movie scripts. In *17th Oriental Chapter of the International Committee for the Co-ordination and Standardization of Speech Databases and Assessment Techniques (COCOSDA)*, pp. 1–4. IEEE, 2014.

[20] Rafael E. Banchs and Haizhou Li. Iris: A chat-oriented dialogue system based on the vector space model. In *Proceedings of the ACL 2012 System Demonstrations*, pp. 37–42, 2012.

[21] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In *proc. EMNLP*, 2014.

[22] Tomas Mikolov, Wen tau Yih, and Geoffrey Zweig. Linguistic regularities in continuous space word representations. In *Proc. NAACL-HLT*, 2013.

[23] M. Schuster and K.K. Paliwal. Bidirectional recurrent neural networks. *Trans. Sig. Proc.*, Vol. 45, pp. 2673–2681, 1997.

[24] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, Vol. 9, pp. 1735–1780, 1997.

[25] Ming Tan, Bing Xiang, and Bowen Zhou. Lstm-based deep learning models for non-factoid answer selection. *CoRR*, Vol. abs/1511.04108, , 2015.

[26] S. K. Michael Wong and Yiyu Yao. An information-theoretic measure of term specificity. *Journal of the American Society for Information Science*, Vol. 43, No. 1, p. 54, 1992.

[27] Rong En Fan, Kai-Wei Chang, Cho-Jui Hsieh, X.-R. Wang, and Chih-Jen Lin. Liblinear: A library for large linear classification. *Journal of Machine Learning Research*, Vol. 9, pp. 1871–1874, 2008.

[28] Joseph L Fleiss. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, Vol. 76, No. 5, pp. 378–382, 1971.