

概要

近年，ニューラルネットワークを用いる手法が自然言語処理の多くのタスクで成果を上げている．その中に対話のモデルをニューラルネットワークにより構築したニューラル対話モデルがある [1]．ニューラル対話モデルの学習に用いるデータをコーパスと呼ぶ．

コーパス中での出現頻度が低い要素を低頻度語という．要素 (語彙) が増えるとニューラルネットワークの学習速度が落ちるため，コーパス中で重要でない低頻度語を同一の記号 (マルチトークン unk_0) へ置き換えることで学習速度の低下を抑える．しかし，マルチトークン unk_0 となった低頻度語は全て同一の記号となるため，コーパスにおいて本来違う単語が同一単語として扱われる．

そこで本研究では，マルチトークン unk_0 を少数のグループへ分割することにより，応答の精度を向上させる．分割するためのラベルとして品詞情報を用いる．

また，過去の発話を用いた学習も行う．非タスク指向型の対話は，状況によって許容される応答が変化する．過去の発話を含まないデータの場合「うん」等の，応答が必ずしも必要でない発話に対する応答の評価を行いにくい．過去の発話を使用し文脈をより明確にすることで，評価を容易にすることを試みた．

結果として，2 発話学習における精度評価の 評価は提案手法が最良で 0.34(68/200) となった．また，精度 評価内における多様性評価は品詞付加のみの手法 (37 種類) と提案手法 (35 種類) が最良だった．精度 + 評価は品詞付加のみの手法と提案手法が最良であり，品詞付加のみの手法が 0.750(150/200) で提案手法が 0.755(151/200) となった．精度 + 評価内での多様性が最良の手法は低頻度語処理をしない手法 (75 種類) だった．また，過去の発話を付加することで評価のしやすさと厳密性が向上した．

目次

第1章	はじめに	1
第2章	ニューラルネットワーク	3
2.1	RNN	3
2.2	LSTM	4
2.3	seq2seq	6
第3章	関連研究	7
3.1	Copyable Model	7
3.2	シリーズ型質問文への対応	7
3.3	低頻度語の高頻度語への置き換え	8
第4章	提案手法	9
4.1	マルチトークンへの品詞情報付加	9
4.2	過去発話の付与	10
第5章	実験環境	11
5.1	使用 NMT エンジン	11
5.2	使用コーパス	11
5.3	実験方法	11
第6章	実験結果	13
6.1	マルチトークン変換数	13
6.2	出力結果	13
6.3	応答精度評価	14
6.3.1	入力に過去の発話を含まない学習データの出力結果評価	15
6.3.2	入力に過去の発話を含む学習データの出力結果評価	18
6.4	出力の多様性評価	19

6.4.1	, 評価における多様性	19
6.5	入出力両方を低頻度語処理した学習データにおける精度・多様性評価 . .	32
6.5.1	入力に過去の発話を含まない学習データの出力結果評価	32
6.5.2	入力に過去の発話を含む学習データの出力結果評価	34
第7章	考察	35
7.1	評価のしやすさ・厳密性	35
7.2	精度	35
7.3	多様性	37
第8章	おわりに	38
第9章	謝辞	39

目 次

2.1 RNN データ入出力	3
2.2 LSTM 内部構造	5
2.3 seq2seq による応答生成モデル	6

表 目 次

3.1.1 Copyable Model の変換例	7
4.1.1 提案手法の変換例	9
4.2.1 過去の発話付与	10
5.2.1 使用コーパス	11
5.2.2 コーパス用途別データ数	11
5.3.1 各手法の変換例	12
6.1.1 Copyable Model による低頻度語変換数	13
6.1.2 提案手法によるヌルトークン変換数	13
6.2.1 過去の発話を含まないデータにおける各手法の応答出力結果	14
6.2.2 過去の発話を含むデータにおける各手法の応答出力結果	14
6.3.1 評価例	15
6.3.2 精度評価結果 (入力に過去の発話を含まない)	16
6.3.3 出力精度評価例 1(入力に過去の発話を含まない)	16
6.3.4 出力精度評価例 2(入力に過去の発話を含まない)	16
6.3.5 精度評価結果 (入力に過去の発話を含まないが, 評価では文脈に過去の発話があると する)	17
6.3.6 出力精度評価例 1(入力に過去の発話を含まないが, 評価では文脈に過去の 発話があるとする)	17
6.3.7 出力精度評価例 2(入力に過去の発話を含まないが, 評価では文脈に過去の 発話があるとする)	18
6.3.8 精度評価結果 (入力に過去の発話を含む)	18
6.3.9 出力精度評価例 (入力に過去の発話を含む)	19
6.3.10 出力精度評価例 (入力に過去の発話を含む)	19
6.4.1 多様性評価結果 (入力に過去の発話を含まない)	20

6.4.2 低頻度語置き換え無し (入力に過去の発話を含まない)	21
6.4.3 低頻度語全てをヌルトークン化 (入力に過去の発話を含まない)	21
6.4.4 Copyable Model(入力に過去の発話を含まない)	22
6.4.5 品詞置き換え (入力に過去の発話を含まない)	22
6.4.6 提案手法 (入力に過去の発話を含まない)	23
6.4.7 多様性評価結果 (入力に過去の発話を含まないが, 評価では文脈に過去の発話があるとする)	24
6.4.8 低頻度語置き換え無し (入力に過去の発話を含まないが, 評価では文脈に過去の発話があるとする)	25
6.4.9 低頻度語全てをヌルトークン化 (入力に過去の発話を含まないが, 評価では文脈に過去の発話があるとする)	25
6.4.10 Copyable Model(入力に過去の発話を含まないが, 評価では文脈に過去の発話があるとする)	26
6.4.11 品詞置き換え (入力に過去の発話を含まないが, 評価では文脈に過去の発話があるとする)	26
6.4.12 提案手法 (入力に過去の発話を含まないが, 評価では文脈に過去の発話があるとする)	27
6.4.13 多様性評価結果 (入力に過去の発話を含む)	28
6.4.14 低頻度語置き換え無し (入力に過去の発話を含む)	29
6.4.15 低頻度語全てをヌルトークン化 (入力に過去の発話を含む)	29
6.4.16 Copyable Model(入力に過去の発話を含む)	30
6.4.17 品詞置き換え (入力に過去の発話を含む)	30
6.4.18 提案手法 (入力に過去の発話を含む)	31
6.4.19 精度及び多様性の増減	32
6.5.1 精度評価結果 (入力に過去の発話を含まない)	33
6.5.2 多様性評価結果 (入力に過去の発話を含まない)	33
6.5.3 精度評価結果 (入力に過去の発話を含まないが, 評価では文脈に過去の発話があるとする)	33
6.5.4 多様性評価結果 (入力に過去の発話を含まないが, 評価では文脈に過去の発話があるとする)	34
6.5.5 精度評価結果 (入力に過去の発話を含む)	34
6.5.6 多様性評価結果 (入力に過去の発話を含む)	34

7.1.1 評価のしやすさが向上した例	35
7.3.1 提案手法の出力数上位 5 文	37

第1章 はじめに

対話のモデルをニューラルネットワークにより構築したものをニューラル対話モデルという [1]。ニューラル対話モデルの学習に用いるデータをコーパスと呼ぶ。コーパス中での出現頻度が低い要素を低頻度語という。要素 (語彙) が増えるとニューラルネットワークの学習速度が落ちることと、低頻度語はうまく学習を行うことができないため、学習の前に低頻度語を削減する処理を行う。

低頻度語処理の手法として、Copyable Model [2] と低頻度語の高頻度語への置き換え [3] がある。Copyable Model は、入力 (発話) と出力 (応答) の単語に対応関係のない対話モデルではヌルトークン unk_0 が多く出現する。また、低頻度語の高頻度語への置き換えを行う手法もあるが、事前準備として単語の置き換え辞書を作成する必要がある。

本研究では、CopyableModel 処理に加え、CopyableModel で処理できずに残ったヌルトークン unk_0 を少数のグループへ分割することにより、応答の精度と多様性の向上を試みる。分割するためのラベルとして品詞情報を用いた。

また、評価のしやすさの向上のために過去の発話を用いた学習を行う。過去の発話を用いた学習では、発話の直前の対話 (過去の発話) を学習データとして追加することで、評価の際に評価者が文脈を知ることができるようにし評価をより厳密にすることを試みる。評価時だけでなく学習データに過去の発話を追加する理由は、偶然の出力ではなく文脈を学習した出力の評価を行うためである。

本研究では入力側 (原言語) のみ低頻度語処理を行う。通常、低頻度語処理は入力、出力ともに行う。学習時間の削減のために必要なのは出力側 (目的言語) であるが、出力側に低頻度語処理を行うと、出力に未知語が含まれるようになるため評価が困難になる。そのため本研究では学習時間を考えず入力の低頻度語処理のみを行う。

本研究の主な主張点を以下に整理する。

- ヌルトークンを削減し、学習データにより多くの情報を残す。これにより応答の精度・多様性を向上を試みた。全てヌルトークンへ置き換える手法より精度・多様性の両方の数値で上回る手法を発見した。精度 + 評価は品詞付加のみの手法と提案手法が最良であり、品詞付加のみの手法が 0.750(150/200) で提案手法が

0.755(151/200)となった。精度 評価内における多様性評価は品詞付加のみの手法 (37 種類) と提案手法 (35 種類) が最良であり、精度 + 評価内での多様性が最良の手法は低頻度語処理をしない手法 (75 種類) だった。

- 過去の発話を用い、文脈を読み取れるデータを作成することで評価を行いやすくした。評価 200 文中、84 文の文脈が明確化した。

本論文の構成は以下の通りである。第 2 章では、ニューラルネットワークについて述べる。第 3 章では、類似研究の手法について述べる。第 4 章では、本研究の手法について述べる。第 5 章では、実験環境について述べる。第 6 章では、実験結果を述べる。第 7 章では、本実験の考察を述べる。第 8 章では、本実験の簡単なまとめを述べる。

第2章 ニューラルネットワーク

2.1 RNN

再帰型ニューラルネット (RNN) は、内部に有向閉路を持つニューラルネットの総称である。この構造により、RNN は情報を一時的に記憶し振る舞いを動的に変化させることが可能となる。RNN は各時刻 t につき 1 つの入力 x_t を受け取り、同時に 1 つの出力 y_t を返す。出力を計算する際、RNN が過去に受け取った入力に反映される。図 2.1 は RNN を順伝播のネットワークに展開したものである。中間層の入力値 z_t は以下の式で表される。 f は活性化関数である。

$$z_t = f(Ux_t + Wz_{t-1})$$

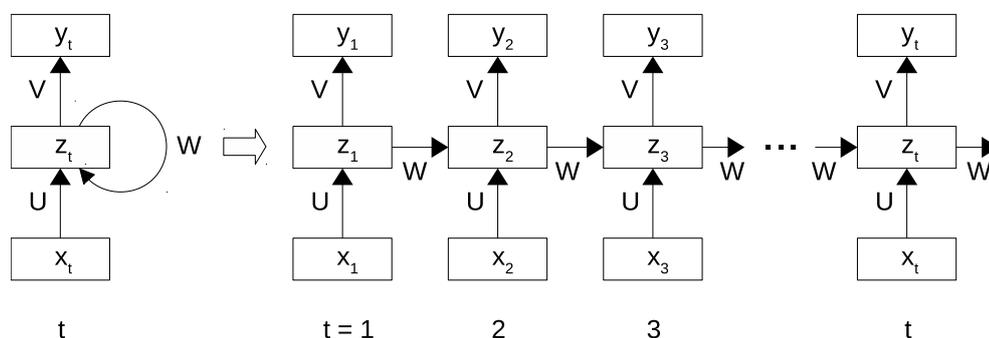


図 2.1: RNN データ入出力

2.2 LSTM

RNNの過去の出力は時刻 t が進んでいく(ニューラルネットの層が深くなる)と、古い過去の出力が反映されなくなる。長期にわたり過去の出力を保持するための方法の一つがLSTMである。図2.2にLSTMの内部構造を示す。中央にメモリセル(図中記号a)がある。メモリセルは状態 s^t を保持し、1時刻を隔ててメモリセル自身に s_t (帰還時は s_{t-1})を帰還させる。周囲に5つのユニット(図中記号b~f)と3つのゲート(入力ゲート、忘却ゲート、出力ゲート)がある。ユニットc, d, fは、 s_t, s_{t-1} にパラメータ(重み)を掛け、外部からの入力 u_I, u_F, u_O と足しあわせ、活性化関数(ロジスティックシグモイド関数) f を適用したものを各ゲートに出力する。ユニットbは前の層からの入力 u_P を受けとり、活性化関数 f を適用する。入力 $u_{P,I,F,O}$ は以下の式で表される。 W_R は帰還値の重みである。

$$u_k = W_n x_t + W_{R,k} z_{t-1} \quad (k = P, I, F, O; n = 1, 2, 3, 4)$$

ゲートはそれぞれゲート値を持つ。ゲート値は活性化関数 f により値域を $[0,1]$ に制約され、この値が1の時にデータが伝達される。各ゲートの出力値は以下の式で表される。

$$g_I = f(u_I + w_I s_{t-1}) = f(W_2 x_t + W_{R,I} z_{t-1} + w_I s_{t-1})$$

$$g_F = f(u_F + w_F s_{t-1}) = f(W_3 x_t + W_{R,F} z_{t-1} + w_F s_{t-1})$$

$$g_O = f(u_O + w_O s_t) = f(W_4 x_t + W_{R,O} z_{t-1} + w_O s_t)$$

ユニットeはメモリセルの出力に活性化関数 f を適用する。状態 s_t と出力 z_t は以下のよう表される[3]。

$$s_t = g_F s_{t-1} + g_I f(u_P)$$

$$z_t = g_O f(s_t)$$

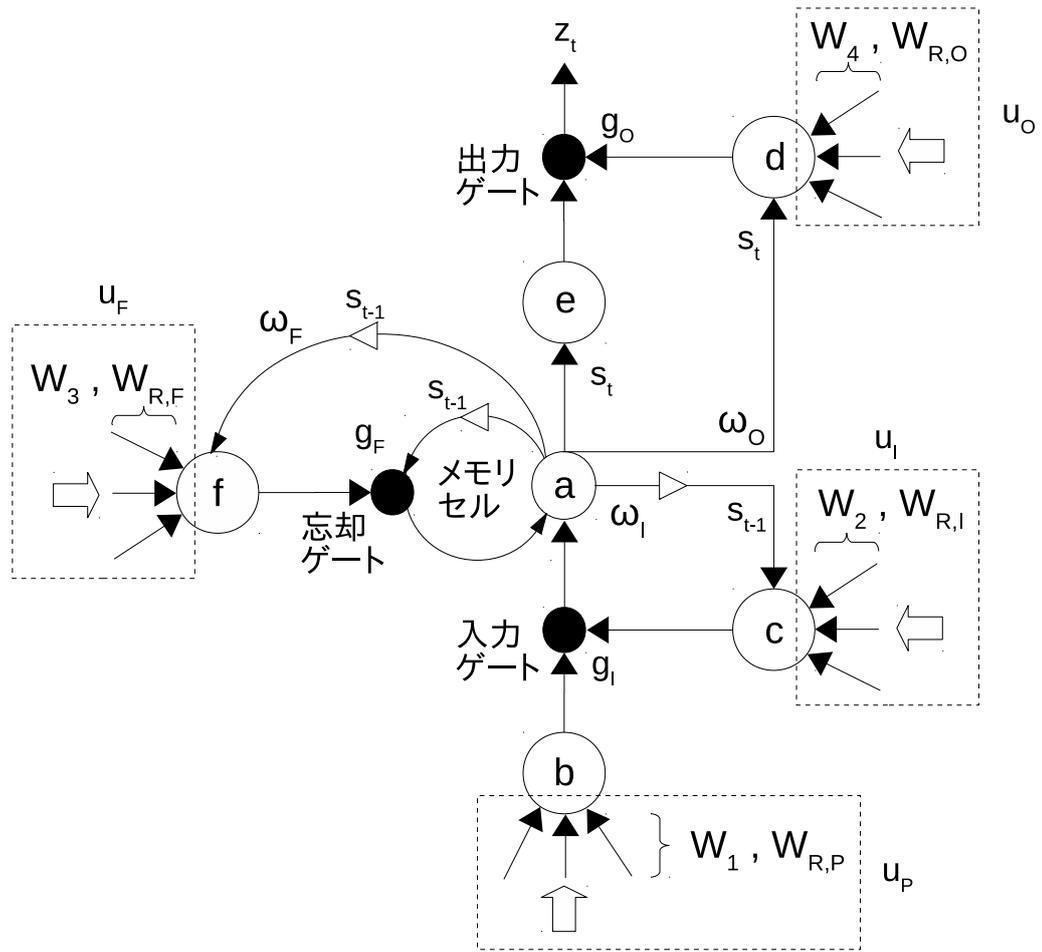


図 2.2: LSTM 内部構造

2.3 seq2seq

seq2seq[5] は NCM モデルのフレームワークである。seq2seq のネットワーク構造を図 2.3 に示す。入力文の各単語が時系列的に入力され、中間層である LSTM ブロックはその時点までの文脈情報を次の中間層へ渡す。通常の RNN では中間層から出力層への出力があるが、seq2seq の入力文側では使用しない。入力文側では最後に文末記号 `<eos>` を読み込み、出力文側の処理へ移る。出力文側では、通常の RNN と同じように、中間層は次の中間層へフィードバックを出力し、同時に出力層への出力を行う。学習時は、出力を教師データと比較して損失を計算する。出力側の入力データは教師データとなる。推論時は、出力は文字列 (単語) の出力となり、出力側の入力データは直前の中間層の出力となる。

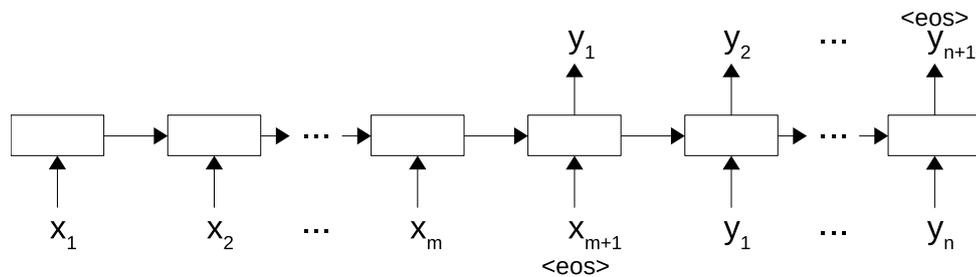


図 2.3: seq2seq による応答生成モデル

第3章 関連研究

3.1 Copyable Model

Copyable Model は NMT のための低頻度語処理モデルである [2] . Copyable Model は unk トークンを一種類だけではなく、複数種類のトークンを使用する . これにより未知語となるヌルトークンに、入力と出力で同じ単語であったという情報を残す .

unk_1, unk_2, unk_3 の順でソース文の低頻度語にナンバーを振る . ターゲット文の未知語アノテーションは、ソース文の未知語化した単語と同じ単語があれば、同じトークンに割り当てる . ターゲット文中の低頻度語に対し、原文に同じ単語が無い場合、対応を持たないヌルトークン unk_0 を使用する . 発話と応答の単語に対応関係のない対話モデルではヌルトークン unk_0 が多く出現する . 表 3.1.1 は CopyableModel の変換例である .

表 3.1.1: Copyable Model の変換例

低頻度語	ソース文	ターゲット文
unk_0 : 5月	unk_0 が待ち遠しいです	こいのぼりを上げるんですか？
unk_0 : ぶら下げ, unk_1 : 蚊取り線香	unk_1 も unk_0 て	unk_1 はほしいですね
unk_1 : 神様, unk_2 : 試験	unk_1 の unk_2 って何よ	unk_1 の unk_2 があるんですか？

3.2 シリーズ型質問文への対応

逐次的になされた質問文を現在の質問文まで連結させて、その連結した文を1つの質問文とし、質問応答処理をする [4] . 現在の質問文以外の質問文での疑問詞はダミー文字に置き換え、質問応答処理で疑問詞としてヒットしないようにする .

本研究では、発話 (過去の発話) とその次に行われた発話を結合し、1つの発話データとして扱う . また、2つの発話間に eos 記号を挿入する .

3.3 低頻度語の高頻度語への置き換え

低頻度語を同義な高頻度語へ置き換える [3] . ターゲット文の , トレーニングデータにおいての低頻度語を高頻度語に言い換えることにより , 未知語への変換を削減する . 単語の置き換え辞書を作成する必要がある .

本研究では , 辞書に依らない置き換えを行う . そのため , Mecab[9] により低頻度語の品詞解析を行い , 未知語を品詞情報付加した未知語へ置き換える .

第4章 提案手法

本研究では，入力側 (原言語) のみにおいて低頻度語処理を行う．これにより，出力に Copyable Model でナンバリングを行ったトークン以外の未知語は出力されなくなり，評価が容易となる．また，未知語が出力されないことにより，同じく未知語が出力されない手法である低頻度語処理を行わない手法との精度の比較を正確に行うことができる．学習時間は考慮に入れず，出力の精度と多様性のみを評価対象とする．

4.1 ヌルトークンへの品詞情報付加

本実験では低頻度語処理の際，Copyable Model の未知語処理を行った後，品詞情報を用いてヌルトークン unk_0 を少数のグループへ分割する．

これにより，低頻度語処理後のコーパスに残すことができる情報が増加する．品詞情報の付加は Mecab[9] で容易に行うことができるため，細分化の方法として採用した．

表 4.1.1 は本研究の手法での低頻度語処理例である．

表 4.1.1: 提案手法の変換例

低頻度語	ソース文	ターゲット文
unk_1 :独自, $unk_{動詞}$:重ねる	unk_1 に $unk_{動詞}$ かもしれない	unk_1 に心の哲学を研究なさっているのですね
$unk_{名詞}$:A(人名), unk_1 :市ヶ谷	$unk_{名詞}$ ね unk_1 なんだよね	unk_1 かー
$unk_{名詞}$:2月11日, unk_1 :建国	$unk_{名詞}$ はなんだろう unk_1 記念か	unk_1 記念か

4.2 過去発話の付与

学習データ, テストデータと評価に過去の発話を付与したデータを使用する実験を行う。今回使用するコーパスにおける対話は「発話 応答(と発話) 応答(と発話)…」となっている。応答は次の応答と組み合わせることで発話へ変化するため、表記としては「応答(と発話)」となる。通常は「発話 応答」のペアに修正し、コーパスとして使用する。

本実験では、発話にその直前の応答を組み合わせ、それを発話データとして用いる。以下に例を示す。

- 発話(直前の応答): 海に行っていないなあ
- 発話(現在の発話): 私も今年はまだ行ってないですね
- 応答: 海は楽しいですね

上記のような対話データを表 4.2.1 のように組み合わせ、コーパスを作成する。eos とは会話の区切りを示す記号であり、本実験ではコーパスに単語として含ませる。

表 4.2.1: 過去の発話付与

入力データ	出力データ
海に行っていないなあ eos 私も今年はまだ行ってないですね	海は楽しいですね

第5章 実験環境

5.1 使用NMTエンジン

本実験には Open-NMT を使用した。バッチサイズは 64, epoch 数は 50 とした。

5.2 使用コーパス

実験で使用するコーパスは, Project Next NLP 対話タスクで収集されたコーパス (雑談対話コーパス)[6] と名大会話コーパス (日本語自然会話書き起こしコーパス)[7] である。コーパスの詳細を表 5.2.1 に示す。

また, 表 5.2.2 はコーパスの用途別データ数である。開発データとは, Open-NMT のプレトレーニングに必要なデータである。

5.3 実験方法

コーパスに出現する低頻度語を, 複数の手法により処理し, その出力の評価を行う。RNN(LSTM) の文生成モデルである seq2seq[5] を使用する。

本研究では, 入力側 (原言語) のみにおいて低頻度語処理を行うが, 比較のために入力, 出力ともに低頻度語処理したモデルの実験も行う。

表 5.2.1: 使用コーパス

コーパス名称	制作
雑談対話コーパス	対話破綻検出チャレンジ
名大会話コーパス	日本語教育ネットワーク

表 5.2.2: コーパス用途別データ数

用途	データ数
総数	69,452
開発データ	1,389
訓練データ	67,369
テストデータ	694

本実験では、過去の発話を含まないデータにおける訓練データの発話(入力側)データ内において一度しか出てこない形態素を低頻度語と定義する。また、入出力両方の低頻度語処理を行う場合、応答(出力側)データの低頻度語は応答データ内において一度しか出てこない形態素を低頻度語と定義する。過去の発話は本来の発話(入力データ)の直前1名分の発話とする。低頻度語数に差を出さないために、過去の発話を含むデータの低頻度語は過去の発話を含まないデータの低頻度語と同一にする。

ニューラルネットワークはOpen-NMTを使用した[8]。バッチサイズは64、epoch数は50とした。

表5.3.1は各手法の低頻度語変換例である。

表 5.3.1: 各手法の変換例

手法	ソース文(発話)	ターゲット文(応答)
原文	電圧を換えるだけで普通に今持っているのが使えるの？	電圧がもし向こうでもオッケーだったら
低頻度語を全てヌルトークンへ変換	unk_0 を unk_0 だけで普通に今持っているのが使えるの？	電圧がもし向こうでもオッケーだったら
Copyable Model	unk_1 を unk_0 だけで普通に今持っているのが使えるの？	unk_1 がもし向こうでもオッケーだったら
提案手法	unk_1 を $unk_{動詞}$ だけで普通に今持っているのが使えるの？	unk_1 がもし向こうでもオッケーだったら

第6章 実験結果

6.1 ヌルトークン変換数

表 6.1.1 は Copyable Model による低頻度語変換数であり，表 6.1.2 は Copyable Model の変換時に発生したヌルトークンに，提案手法により品詞情報を付加して行った分類の変換数である．

表 6.1.1: Copyable Model による低頻度語変換数

unk 番号	変換数
0(ヌルトークン)	8,701
1	105
2	3

表 6.1.2: 提案手法によるヌルトークン変換数

品詞	変換数	品詞	変換数
名詞	6,319	記号	13
動詞	1,508	助詞	10
副詞	390	助動詞	10
形容詞	389	連体詞	8
感動詞	34	接続詞	5
接頭詞	15		

6.2 出力結果

表 6.2.1 は各手法の，過去の発話を含まないデータにおける応答出力結果例である．表 6.2.2 は各手法の，過去の発話を含めた学習データにおける応答出力結果例である．過去の発話は本来の発話（入力データ）の直前 1 名分の発話である．

表 6.2.1: 過去の発話を含まないデータにおける各手法の応答出力結果

発話：寝不足のときは頭のてっぺんを軽く押すといいんだって	
手法	入出力
Copyable Model	発話：寝不足のときは頭の unk_0 を軽く unk_0 といいんだって 応答：[A] さんとね [A] さんが来たの？
提案手法	発話：寝不足のときは頭の $unk_{名詞}$ を軽く $unk_{動詞}$ といいんだって 応答：そうそう

表 6.2.2: 過去の発話を含むデータにおける各手法の応答出力結果

発話：歯医者の日なんや eos 歯科衛生士さんは美人揃いでした	
手法	入出力
Copyable Model	発話：歯医者の日なんや eos unk_0 さんは美人 unk_0 でした 応答：そうそう そうそう
提案手法	発話：歯医者の日なんや eos $unk_{名詞}$ さんは美人 $unk_{名詞}$ でした 応答：あ そうなんだ

6.3 応答精度評価

評価は各手法の出力をテストデータから抽出した 200 文で行った。今回の実験では、過去の発話を含むデータ、含まないデータの両方とも、入力より前の文脈(データ)を見ずに評価し、必要であれば入力より前の文脈を出力に都合の良いように補完する。大きく分けて、 \circ は極端な補完無く評価可能且つ適切な応答、 \ominus は一部不適切だが文脈に沿う意図が読み取れる応答、 \times は不適切な応答であるが、それだけでの評価は困難であるため、更に詳細な評価基準を定めた。また、評価例を表 6.3.1 に示す。

- 評価：
 - － 現状の発話応答で正しく完結している。
 - － 意志表示が必要な発話に対し、自分の意志表示を行う相槌。
- 評価：
 - － 理解可能だが文(文法)が一部破綻。

- 意図が理解可能な範囲で，期待する内容からずれた返答．
- より前の対話があると仮定すれば (極端な文脈を補完すれば) 正しいと思われる
 応答．
- 応答として適切でないが明らかにジョークと取れる応答．
- 応答として適切だが会話の進展が無い．
- 会話として適当なオウム返し．
- 相手に発言を促すだけの相槌．
- 質問に対する聞き返し．
- 返答として適切と思われるが一文で完結していない発話．

● 評価：×

- 理解可能だが文が完全に破綻．
- 理解不可能．
- 会話として適当で無いオウム返し，相槌．

表 6.3.1: 評価例

評価理由	入出力
: 現状の発話応答で正しく完結している	発話：こんにちは 夏 といえば スイカ だね 応答：スイカ は 大好き です ね
: 応答として適切だが会話の進展が無い	発話：奥 が 深い です ね 応答：はい
: 意図が理解可能な範囲で期待する内容 からずれた返答	発話：スポーツ は なさい ます か？ 応答：スポーツ は 必要 です
: 会話として適当なオウム返し	発話：マスクメロン は 美味しい です ね 応答：美味しい です ね
×: 会話として適当で無いオウム返し，相 槌	発話：沖縄 いつ 行く の？ 応答：沖縄

6.3.1 入力に過去の発話を含まない学習データの出力結果評価

6.3.1.1 評価に過去の発話を使用しない

入力に過去の発話を含まないデータで学習したモデルにおける，テスト出力結果の精度評価を表 6.3.2 に示す．また，出力の精度評価例を表 6.3.3 と表 6.3.4 に示す．

表 6.3.2: 精度評価結果 (入力に過去の発話を含まない)

手法			×
低頻度語置き換え無し	0.310(62/200)	0.455(91/200)	0.235(47/200)
低頻度語全てをヌルトークン化	0.380(76/200)	0.430(86/200)	0.190(38/200)
Copyable Model	0.345(69/200)	0.410(82/200)	0.245(49/200)
品詞置き換え	0.300(60/200)	0.515(103/200)	0.185(37/200)
提案手法	0.360(72/200)	0.445(89/200)	0.195(39/200)

表 6.3.3: 出力精度評価例 1(入力に過去の発話を含まない)

手法	入出力	評価
低頻度語置き換え無し	発話：儀式は必要だそうですよ 応答：勇気は必要ですね	×
低頻度語全てをヌルトークン化	発話： unk_0 は必要だそうですよ 応答：そうですね	
Copyable Model	発話： unk_1 は必要だそうですよ 応答： unk_1 が多いです	
品詞置き換え	発話： $unk_{名詞}$ は必要だそうですよ 応答：そうだよね	
提案手法	発話： unk_1 は必要だそうですよ 応答： unk_1 がいいですね	×

表 6.3.4: 出力精度評価例 2(入力に過去の発話を含まない)

手法	入出力	評価
低頻度語置き換え無し	発話：私 1 限 全部 外したよ 応答：ほんと	
低頻度語全てをヌルトークン化	発話：私 1 限 全部 外したよ 応答：そう	
Copyable Model	発話：私 1 限 全部 外したよ 応答：そう	
品詞置き換え	発話：私 1 限 全部 外したよ 応答：あー	
提案手法	発話：私 1 限 全部 外したよ 応答：あっ そう	

6.3.1.2 評価に過去の発話を使用する

学習・テスト時に過去の発話を使用しなかったモデルの出力(応答)を，過去の発話を含むモデルと同様な評価方法により評価した．すなわち，本来の発話データとその直前一発話分の発話を文脈とし，それ以前の文脈は見ずに評価する．精度評価を表 6.3.5 に示

す。また，出力の精度評価例を表 6.3.6，表 6.3.7 に示す。学習時・テスト時は，実際には過去発話は使用しておらず，評価にのみ過去発話を使用する。

表 6.3.5: 精度評価結果 (入力に過去の発話を含まないが，評価では文脈に過去の発話があるとする)

手法			×
低頻度語置き換え無し	0.280(56/200)	0.420(84/200)	0.300(60/200)
低頻度語全てをヌルトークン化	0.370(74/200)	0.385(77/200)	0.245(49/200)
Copyable Model	0.290(58/200)	0.400(80/200)	0.310(62/200)
品詞置き換え	0.285(57/200)	0.415(83/200)	0.300(60/200)
提案手法	0.345(69/200)	0.400(80/200)	0.255(51/200)

表 6.3.6: 出力精度評価例 1(入力に過去の発話を含まないが，評価では文脈に過去の発話があるとする)

手法	入出力	評価
低頻度語置き換え無し	発話：(何かの儀式ですか) eos 儀式は必要だそうですよ 応答：勇気は必要ですね	
低頻度語全てをヌルトークン化	発話：(何かの儀式ですか) eos unk_0 は必要だそうですよ 応答：そうですね	
Copyable Model	発話：(何かの儀式ですか) eos unk_1 は必要だそうですよ 応答： unk_1 が多いです	
品詞置き換え	発話：(何かの儀式ですか) eos $unk_{名詞}$ は必要だそうですよ 応答：そうだよね	
提案手法	発話：(何かの儀式ですか) eos unk_1 は必要だそうですよ 応答： unk_1 がいいですね	×

表 6.3.7: 出力精度評価例 2(入力に過去の発話を含まないが, 評価では文脈に過去の発話があるとする)

手法	入出力	評価
低頻度語置き換え無し	発話:(あー私それより明日の1限が心配) eos 私1限全部外したよ 応答:ほんと	
低頻度語全てをヌルトークン化	発話:(あー私それより明日の1限が心配) eos 私1限全部外したよ 応答:そう	
Copyable Model	発話:(あー私それより明日の1限が心配) eos 私1限全部外したよ 応答:そう	
品詞置き換え	発話:(あー私それより明日の1限が心配) eos 私1限全部外したよ 応答:あー	
提案手法	発話:(あー私それより明日の1限が心配) eos 私1限全部外したよ 応答:あっそう	

6.3.2 入力に過去の発話を含む学習データの出力結果評価

入力に過去の発話を含んだデータで学習したモデルにおける, テスト出力結果の精度評価を表 6.3.8 に示す. また, 出力の精度評価例を表 6.3.9, 表 6.3.10 に示す.

表 6.3.8: 精度評価結果(入力に過去の発話を含む)

手法			×
低頻度語置き換え無し	0.270(54/200)	0.430(86/200)	0.300(60/200)
低頻度語全てをヌルトークン化	0.315(63/200)	0.385(77/200)	0.300(60/200)
Copyable Model	0.325(65/200)	0.375(75/200)	0.300(60/200)
品詞置き換え	0.290(58/200)	0.460(92/200)	0.250(50/200)
提案手法	0.340(68/200)	0.415(83/200)	0.245(49/200)

表 6.3.9: 出力精度評価例 (入力に過去の発話を含む)

手法	入出力	評価
低頻度語置き換え 無し	発話：何かの儀式ですか eos 儀式は必要だそうですよ 応答：そうですね	
低頻度語全てをヌ ルトークン化	発話：何かの unk_0 ですか eos unk_0 は必要だそうですよ 応答：そうですね	
Copyable Model	発話：何かの unk_1 ですか eos unk_1 は必要だそうですよ 応答： unk_1 が多いです	
品詞置き換え	発話：何かの $unk_{名詞}$ ですか eos $unk_{名詞}$ は必要だそうですよ 応答：そうですか	
提案手法	発話：何かの unk_1 ですか eos unk_1 は必要だそうですよ 応答： unk_1 が多いです	

表 6.3.10: 出力精度評価例 (入力に過去の発話を含む)

手法	入出力	評価
低頻度語置き換え 無し	発話：あー私それより明日の1限が心配 eos 私1限全部外したよ 応答：ありがとうございます	
低頻度語全てをヌ ルトークン化	発話：あー私それより明日の1限が心配 eos 私1限全部外したよ 応答：大丈夫だね	
Copyable Model	発話：あー私それより明日の1限が心配 eos 私1限全部外したよ 応答：いいです	×
品詞置き換え	発話：あー私それより明日の1限が心配 eos 私1限全部外したよ 応答：ありがとうございます	
提案手法	発話：あー私それより明日の1限が心配 eos 私1限全部外したよ 応答：はい	

6.4 出力の多様性評価

6.4.1 , 評価における多様性

評価は精度評価と同じ出力で行った。精度評価で 及び 評価だったものの、出力文の種類数を計算した。

6.4.1.1 過去の発話を含まない学習データの多様性評価

入力に過去の発話を含まないデータで学習したモデルにおける，テスト出力結果の多様性評価を表 6.4.1 に示す．また，各手法の 評価においての出力と出力数詳細を表 6.4.2 から表 6.4.6 に示す．

表 6.4.1: 多様性評価結果 (入力に過去の発話を含まない)

手法	評価数	評価数	出力種類数	出力種類数
低頻度語置き換え無し	62	91	34	31
低頻度語全てをヌルトークン化	76	86	39	25
Copyable Model	69	82	38	22
品詞置き換え	60	103	30	32
提案手法	72	89	42	25

表 6.4.2: 低頻度語置き換え無し (入力に過去の発話を含まない)

出力数	応答
26	うん
2	海に行きませんか？
2	ねー
2	そうなんですか
1	僕も大好きです
1	美味しいですね
1	行きたい
1	気持ちがいいですね
1	気をつけてください
1	楽しいですね
1	確かにね
1	何が好きですか？
1	一人で行くのです
1	ほんと？
1	ほんと
1	へえー
1	はい
1	どういたしまして
1	そうですね
1	そうそう そうそう
1	そうなんですか？
1	そうなんだ
1	そうそう そうそう
1	そう
1	かわいい
1	ええ
1	うん
1	いいですね
1	ありがとう
1	あっ そう
1	ああ そうですか
1	ああ そう
1	あ そう
1	いい？

表 6.4.3: 低頻度語全てをマルチトークン化 (入力に過去の発話を含まない)

出力数	応答
28	うん
5	あ そうなんだ
4	そうですね
3	そうなんだ
2	そうなんですか
1	和食が好きですね
1	予防が大切ですね
1	味がいいですね
1	湘南は楽しいですね
1	私も好きです
1	嫌いです
1	気持ちがいいですね
1	楽しいですね
1	海に行きませんか？
1	何？
1	穏やかだね
1	一人で行くのです
1	ベストがいいですね
1	スポーツは得意ですか？
1	ほんと？
1	へえー
1	はい
1	どういたしまして
1	そうですね
1	そうそう そうそう
1	そう そうそう
1	そう
1	かわいいね
1	おいしい
1	ええ
1	え？
1	うそー
1	いきたい
1	いいなー
1	いいですね
1	ありがとう
1	あほんと
1	あ そうなんですか
1	そうですね

表 6.4.4: Copyable Model(入力に過去の発話を含まない)

出力数	応答
27	うん
5	そうですね
2	あ そうなんだ
1	麦茶が欲しいですね
1	川は楽しいですね
1	湘南は楽しいですね
1	行きたい
1	好きですか？
1	気持ちがいいですね
1	楽しいですね
1	海に行きたいですね
1	何を？
1	何？
1	一人で行くのです
1	スポーツは好きですか？
1	ほんと？
1	はあ
1	どういたしまして
1	そうそう そうそう
1	そうなんだ
1	そうなのですか？
1	そうだね
1	そう
1	おいしい
1	ええ
1	うーん
1	いや
1	いいよ
1	いいですね
1	いいですね
1	ありがとう
1	あっ そう
1	あ 本当？
1	あ ほんと
1	あ そう
1	<i>unk₁</i> はいいですね
1	<i>unk₁</i> ?
1	<i>unk₁</i> が多いです

表 6.4.5: 品詞置き換え (入力に過去の発話を含まない)

出力数	応答
24	うん
5	そうですね
3	そうなんだ
2	そうそう そうそう
1	私も好きです
1	気持ちがいいですね
1	気をつけてください
1	楽しいですね
1	海は楽しいですね
1	音楽は好きですね
1	一人で行くのです
1	テレビは好きですか？
1	わかんない
1	ほんと
1	なんで？
1	どういたしまして
1	そうなんですか？
1	そうなんですか
1	そうなのですか
1	そうだね
1	そうそうそう
1	かわいいね
1	うんうん
1	いいですね
1	あった？
1	あ そうなんだ
1	あ そうなの？
1	あ そうですか
1	あ そう
1	そうだよ

表 6.4.6: 提案手法 (入力に過去の発話を含まない)

出力数	応答
23	うん
2	そうですね
2	そうそう そうそう
2	そう なんだ
2	そうですか
2	うん?
2	あっ そう
2	あ ほんと
2	あ そうですか
1	予防が大切ですね
1	川は楽しいですね
1	食べない
1	松たか子さんも好きですか?
1	行きたい
1	汗で水分が出ますよねー
1	楽しいですね
1	海に行きませんか?
1	何が好きですか?
1	何が?
1	塩が好きなんですか?
1	一人で行くのです
1	ベストがいいですね
1	やってた
1	ほんと?
1	はい
1	なんで?
1	どういたしまして
1	そうそう
1	そうなんですか
1	そうですよね
1	そうだね
1	おいしいですね
1	え?
1	いや
1	いきたーい
1	いいですね
1	ありがとう
1	ありがとう
1	あ そう
1	unk_1 は好きですね
1	unk_1 は好きですか?
1	何で?

6.4.1.2 過去の発話を含まない学習データの過去の発話を用いた多様性評価

過去の発話を含むモデルと同様な評価方法により評価した過去の発話を含まない学習モデルにおけるテスト出力結果の多様性評価を表 6.4.7 に示す。また、各手法の 評価においての出力と出力数詳細を表 6.4.8 から表 6.4.12 に示す。

表 6.4.7: 多様性評価結果 (入力に過去の発話を含まないが、評価では文脈に過去の発話があるとする)

手法	評価数	評価数	出力種類数	出力種類数
低頻度語置き換え無し	56	84	33	30
低頻度語全てをヌルトークン化	74	77	39	24
Copyable Model	58	80	29	27
品詞置き換え	57	83	26	29
提案手法	69	80	41	22

表 6.4.8: 低頻度語置き換え無し (入力に過去の発話を含まないが, 評価では文脈に過去の発話があるとする)

出力数	応答
22	うん
2	はい
2	そうなんですか
1	僕も大好きです
1	美味しいですね
1	大変だよ
1	川は綺麗ですね
1	行きたい
1	気持ちがいいですね
1	気をつけてください
1	楽しいですね
1	海に行きませんか?
1	何が好きですか?
1	一人で行くの
1	わからん
1	ほんと?
1	ほんと
1	ねー
1	どういたしまして
1	そうそう そうそう
1	そうなんだ
1	そう そうそう そうそう
1	そう そう そう そう
1	そう
1	かわいい
1	ええ
1	うん
1	いいですね
1	あっ そう
1	ああ そうですか
1	ああ そう
1	あ そう
1	いい?

表 6.4.9: 低頻度語全てをヌルトークン化 (入力に過去の発話を含まないが, 評価では文脈に過去の発話があるとする)

出力数	応答
28	うん
4	あ そうなんだ
3	そうなんだ
2	はい
2	そうですね
2	いいですね
1	和食が好きですね
1	予防が大切ですね
1	味がいいですね
1	川は綺麗ですね
1	湘南は楽しいですね
1	私も好きです
1	嫌いなんですか?
1	嫌いです
1	気持ちがいいですね
1	楽しいですね
1	海に行きませんか?
1	何?
1	穏やかだね
1	一人で行くの
1	テレビで見ます?
1	スポーツは得意ですか?
1	ほんと?
1	へえー
1	どういたしまして
1	そうですね
1	そうそう そうそう
1	そうなんですか
1	そうだよ
1	そう そうそう
1	そう
1	かわいいね
1	おいしい
1	ええ
1	うそー
1	いきたい
1	あほんと
1	あ そうなんですか
1	あ そうなんだ

表 6.4.10: Copyable Model(入力に過去の発話を含まないが, 評価では文脈に過去の発話があるとする)

出力数	応答
26	うん
3	そうですね
2	そう
2	あ そうなんだ
1	湘南は楽しいですね
1	行きたい
1	好きですか?
1	気持ちがいいですね
1	楽しいですね
1	海は楽しいですね
1	何を?
1	一人で行くのです
1	スポーツは好きですか?
1	まじで?
1	ほんと?
1	はあ
1	どういたしまして
1	そうそう そうそう
1	そうなんだ
1	そうだね
1	おいしい
1	ええ
1	いや
1	あっ そう
1	あ 本当?
1	あ ほんと
1	あ そう
1	<i>unk₁</i> はいいですね
1	<i>unk₁</i> が多いです

表 6.4.11: 品詞置き換え (入力に過去の発話を含まないが, 評価では文脈に過去の発話があるとする)

出力数	応答
23	うん
5	そうですね
3	そうなんだ
2	海は楽しいですね
2	そうそう そうそう
2	そう そうそう
1	気持ちがいいですね
1	気をつけてください
1	楽しいですね
1	一人で行くのです
1	テレビは好きですか?
1	わかんない
1	はい
1	どういたしまして
1	そうそう
1	そうなんですか
1	そうなのですか
1	そうだね
1	かわいいね
1	うん
1	いいですね
1	あった?
1	あ そうなんだ
1	あ そうなの?
1	あ そうですか
1	そうだよ

表 6.4.12: 提案手法 (入力に過去の発話を含まないが, 評価では文脈に過去の発話があるとする)

出力数	応答
23	うん
2	はい
2	そうですね
2	そうそう そうそう
2	あっ そう
2	あほんと
2	あ そうですね
1	予防が大切ですね
1	食べない
1	湘南がいいですね
1	行きたい
1	汗で水分が出ますよねー
1	楽しいですね
1	海は穏やかですね
1	海に行きませんか?
1	何が?
1	塩が好きなんですか?
1	一人で行くのです
1	わからん
1	やってた
1	ほんと?
1	なんで?
1	どういたしまして
1	そうそう
1	そうなんですか
1	そうなんだ
1	そうですね
1	そうですね
1	そうだね
1	そう
1	おいしいですね
1	うんうん
1	うん?
1	うん
1	いや
1	いきたーい
1	いいですね
1	あ そう
1	unk_1 は好きですね
1	unk_1 は好きですか?
1	何で?

6.4.1.3 過去の発話を含む学習データの出力結果評価

入力に過去の発話を含んだデータで学習したモデルにおける，テスト出力結果の多様性評価を表 6.3.8 に示す．また，各手法の 評価においての出力と出力数詳細を表 6.4.14 から表 6.4.18 に示す．表内の [A] は人名 (固有名詞) を記号へ置き換えたものである．

表 6.4.13: 多様性評価結果 (入力に過去の発話を含む)

手法	評価数	評価数	出力種類数	出力種類数
低頻度語置き換え無し	54	86	31	44
低頻度語全てをヌルトークン化	63	77	30	24
Copyable Model	65	75	27	20
品詞置き換え	58	92	37	29
提案手法	68	83	35	27

表 6.4.14: 低頻度語置き換え無し (入力に過去の発話を含む)

出力数	応答
20	うん
3	そうなんだ
2	なんで?
2	そうなんですか
1	美味しいですね
1	私も好きです
1	気持ちがいいですね
1	海に行きたいですね
1	一人で行ってないです
1	サーフィンは楽しいですね
1	ほんとー
1	はい
1	どんな仕事が好きですか?
1	そうですね
1	そうそう そうそう
1	そうそう
1	そうなんですか
? 1	そうなの?
1	そうですか
1	そう そうそう
1	そう
1	かわいいね
1	うんうん
1	うん そう そうそう
1	うん うん うん うん
1	うそー
1	いいですねー
1	いいじゃん
1	あっほんと?
1	あっ そうなんだ
1	あうそー

表 6.4.15: 低頻度語全てをヌルトークン化 (入力に過去の発話を含む)

出力数	応答
28	うん
5	そうですね
2	ほんと?
2	そうなんだ
1	味がいいですね
1	怖いよねー
1	私も大好きです
1	行こう
1	楽しいですね
1	海に行きたいですね
1	サーフィンは楽しいですね
1	わかんない
1	はい
1	なんで?
1	どういたしまして
1	それは分かりません
1	それはそうなんだ
1	そう思います
1	そうなんだ そうですね
1	そうですか
1	そう そうそう
1	え?
1	いただきまーす
1	いいですね
1	いいじゃん
1	あほんと
1	あ そうなんだ
1	あ そう
1	[A] ちゃんのやつ?
1	あ そうなの

表 6.4.16: Copyable Model(入力に過去の発話を含む)

出力数	応答
29	うん
5	そうですね
2	気持ちがいいですね
2	海は楽しいですね
2	そうですか
2	そうだよ
2	あ そうなんだ
2	<i>unk</i> ₁ が多いですね
1	大丈夫
1	楽しいよ
1	わかんない
1	まじで?
1	なんで?
1	なんで?
1	なるほど
1	そうそう そうそう
1	そうなんですか
1	そうなの?
1	そうですか?
1	そうだね
1	え?
1	いいよね
1	いいの?
1	あった
1	あ そう
1	<i>unk</i> ₁ は好きですね
1	<i>unk</i> ₁ が多いですね

表 6.4.17: 品詞置き換え (入力に過去の発話を含む)

出力数	応答
17	うん
5	そうですね
2	私も好きです
1	味が好きですね
1	朝から海に行きますか?
1	知らない
1	行こう
1	気持ちがいいですね
1	楽しいよ
1	何が?
1	何?
1	まじで?
1	なんで?
1	なるほど
1	そんなことない
1	それはないよね
1	そうですね
1	そうそう そうそう
1	そうなんですよ
1	そうなんだよね
1	そうなんだ
1	そうですね
1	そうですか
1	そうだね
1	そう
1	かわいいね
1	お酒は美味しいですね
1	ええ
1	ううん
1	いいよ
1	あっ そうなんだ
1	あっ そうなの
1	あー そっか そっか
1	あほんと
1	あ そうなんですか
1	[A]さん
1	そうですか

表 6.4.18: 提案手法 (入力に過去の発話を含む)

出力数	応答
27	うん
3	そうですね
3	そうなんだ
2	確かに
2	何？
2	そうそう そうそう
1	怖いねー
1	微妙だよ
1	楽しいよ
1	海に行きたいですね
1	一人で行くのです
1	サーフィンは楽しいですね
1	なんで？
1	どんな料理が好きですか？
1	そうなんですか？
1	そうですか
1	そうだねー
1	そうだね
1	そう そうそう そうそう
1	これもいい
1	ええ
1	うん
1	うーん そうですね
1	いいよ
1	いいですね
1	あった
1	あっ そうなんだ
1	ああ そうなんだ
1	あー そうなんだ
1	あ 本当？
1	あ 本当
1	あ そうなんですか
1	unk_1 は美味しいですね
1	unk_1 が多いですね
1	unk_1 が多いですね

6.4.1.4 各手法における多様性の増減

表 6.4.19 は過去の発話を含むモデルと同様な評価方法により評価した過去の発話を含まない学習モデルと入力に過去の発話を含んだデータで学習したモデルの比較による増減である。

表 6.4.19: 精度及び多様性の増減

手法	評価数	評価数	出力種類数	出力種類数
低頻度語置き換え無し	-2	+2	-2	+14
低頻度語全てをヌルトークン化	-11	0	-9	0
Copyable Model	+7	-5	-2	-7
品詞置き換え	+1	+9	+11	0
提案手法	-1	+3	-6	+5

6.5 入出力両方を低頻度語処理した学習データにおける精度・多様性評価

以下に入出力の両方に低頻度語処理を行った学習データを使用したモデルによる出力の、各手法の精度・多様性評価を記載する。また、評価を発話(入力)のみ低頻度語処理した学習データを使用したモデルによる出力と比較した。精度評価の基準は、発話(入力)のみ低頻度語処理したデータで使用したものに以下の条件を加える。

- 評価：
 - － 他の 評価条件に加え，Copyable Model でナンバリングしたトークン以外の未知語が含まれない
- 評価：
 - － 文脈から未知語を補完でき，且つ理解可能である
- 評価：×
 - － 未知語がわからなくても理解可能だが，未知語の補完が不可能
 - － 未知語の補完が可能だが，返答として理解不可能

6.5.1 入力に過去の発話を含まない学習データの出力結果評価

6.5.1.1 評価に過去の発話を使用しない

入力に過去の発話を含まないデータで学習したモデルにおける，テスト出力結果の精度評価を表 6.5.1 に，多様性評価を表 6.5.2 に示す。表 6.5.1 の括弧内右側の数値は，評価

を発話のみ低頻度語処理した学習データを使用したモデルによる出力との精度の比較である。

表 6.5.1: 精度評価結果 (入力に過去の発話を含まない)

手法	評価数	評価数	× 評価数
低頻度語全てを ヌルトークン化	0.335(67/200,-0.045)	0.470(94/200,+0.04)	0.195(39/200,+0.005)
Copyable Model	0.285(57/200,-0.060)	0.485(97/200,+0.075)	0.230(46/200,-0.015)
品詞置き換え	0.305(61/200,+0.005)	0.505(101/200,-0.010)	0.190(38/200,+0.005)
提案手法	0.355(71/200,-0.005)	0.410(82/200,-0.035)	0.235(47/200,+0.040)

表 6.5.2: 多様性評価結果 (入力に過去の発話を含まない)

手法	評価数	評価数	種類数	種類数
全ヌルトークン化	67	94	31	32
Copyable Model	57	97	27	39
品詞置き換え	61	101	28	42
提案手法	71	82	36	27

6.5.1.2 評価に過去の発話を使用する

学習・テスト時に過去の発話を使用しなかったモデルの出力(応答)を、過去の発話を含むモデルと同様な評価方法により評価した。すなわち、本来の発話データとその直前一発話分の発話を文脈とし、それ以前の文脈は見ずに評価する。精度評価を表 6.5.3 に、多様性評価を表 6.5.4 に示す。表 6.5.3 の括弧内右側の数値は、評価を発話のみ低頻度語処理した学習データを使用したモデルによる出力との精度の比較である。

表 6.5.3: 精度評価結果 (入力に過去の発話を含まないが、評価では文脈に過去の発話があるとする)

手法	評価数	評価数	× 評価数
低頻度語全てを ヌルトークン化	0.360(72/200,-0.010)	0.345(69/200,-0.040)	0.295(59/200,+0.050)
Copyable Model	0.260(52/200,-0.030)	0.425(85/200,+0.025)	0.315(63/200,+0.005)
品詞置き換え	0.260(52/200,-0.025)	0.445(89/200,+0.030)	0.295(59/200,-0.005)
提案手法	0.315(63/200,-0.030)	0.395(79/200,-0.005)	0.290(58/200,+0.035)

表 6.5.4: 多様性評価結果 (入力に過去の発話を含まないが, 評価では文脈に過去の発話があるとする)

手法	評価数	評価数	種類数	種類数
全ヌルトークン化	72	69	29	27
Copyable Model	52	85	26	37
品詞置き換え	52	89	25	36
提案手法	63	79	33	27

6.5.2 入力に過去の発話を含む学習データの出力結果評価

入力に過去の発話を含んだデータで学習したモデルにおける, テスト出力結果の精度評価を表 6.5.5 に, 多様性評価を表 6.5.6 に示す. 表 6.5.5 の括弧内右側の数値は, 評価を発話のみ低頻度語処理した学習データを使用したモデルによる出力との精度の比較である.

表 6.5.5: 精度評価結果 (入力に過去の発話を含む)

手法	評価数	評価数	× 評価数
低頻度語全てをヌルトークン化	0.250(50/200, -0.065)	0.425(85/200, +0.040)	0.325(65/200, +0.025)
Copyable Model	0.260(52/200, -0.065)	0.430(86/200, +0.055)	0.310(62/200, +0.010)
品詞置き換え	0.265(53/200, -0.025)	0.410(82/200, -0.050)	0.325(65/200, +0.075)
提案手法	0.250(50/200, -0.090)	0.445(89/200, +0.030)	0.305(61/200, +0.060)

表 6.5.6: 多様性評価結果 (入力に過去の発話を含む)

手法	評価数	評価数	種類数	種類数
全ヌルトークン化	50	85	23	34
Copyable Model	52	86	21	31
品詞置き換え	53	82	23	31
提案手法	50	89	24	36

第7章 考察

7.1 評価のしやすさ・厳密性

入力に過去の発話を含むデータで評価することで、文脈がより分かりやすくなった。評価 200 文中、84 文の文脈が明確化した。残り 116 文は、現在の発話のみで文脈が明確なもの、明確化のために更に過去の発話が必要な文と、現在の発話と過去の発話で話題の転換が起こった文である。また、応答を更に厳密に評価することが可能となった。入力に過去の発話を含まない学習データにおいて、過去の発話を含むモデルと同様の評価を行った場合、全ての手法において精度が低下したのは厳密性が向上したためである。表 7.1.1 は評価のしやすさが向上した発話と応答の例である。

表 7.1.1: 評価のしやすさが向上した例

発話	応答
前期やった? eos スペイン語で	まじで?
理由:前期の外国語の授業でスペイン語を選択したという発話だったことがわかる	
何時待ち合わせ? eos 7時前くらい	あ そうなんだ
理由:待ち合わせ時間を聞いていたことがわかる	

また、応答(出力)の低頻度語処理をしないことにより、応答に未知語が含まれなくなるため評価のしやすさが向上するとともに、低頻度語置き換え無しの手法との同条件での比較が可能となった。応答に未知語が含まれる場合、未知語が含まれない低頻度語置き換え無しの手法が最も精度がよくなる。応答の低頻度語処理をしないことにより、低頻度語置き換え無しの手法と同じく未知語が含まれない出力で比較することが可能となった。

7.2 精度

入力に過去の発話を含まない学習データと過去の発話を含む学習データの出力比較では、過去の発話を含む学習データの出力において品詞置き換えのみの精度評価が 評価で+1 向上し、 評価で+9 向上した(表 6.4.19 参照)。入力に過去の発話を含む学習デー

タの 評価においては提案手法が 0.34(68/200) となり最良だった (表 6.3.8 参照) . 入力に過去の発話を含む学習データでは , 精度が大きく減少したのは低頻度語を全てヌルトークン化する手法のみだった (表 6.4.19 参照) .

過去の発話を付与すると入力側の学習データが長くなるため , 精度が落ちると思われたが , 品詞付加の手法では精度が向上し , Copyable Model と提案手法では精度の低下は無かったため , 精度を維持したまま過去の発話を学習データとして使用できる可能性がある .

また , 過去の発話を用いた評価の場合 , 入出力の両方に低頻度語処理を行った結果と入力だけに低頻度語処理を行った結果を比べると , 両方に処理を行った結果の 評価が全ての手法において減少した (表 6.5.1 , 表 6.5.3 , 表 6.5.5 参照) . 応答に未知語が出力された場合 , 未知語が含まれない場合と比べて評価と評価のしやすさが低くなる . また , 低頻度語処理を行わないモデルは未知語が出力されないため , 出力側で低頻度語処理を行うモデルより性能が高く評価されてしまう . そのため , 未知語が出力されない手法と正しく比較するためにも , 対話の評価においては低頻度語処理を入力側のみ行うことが良いと思われる .

7.3 多様性

入力に過去の発話を含む学習データにおける精度 評価の多様性は、品詞置き換えのみの手法が 37 種類で最良であり (表 6.4.13 参照)、過去の発話を含まない学習データからの向上数も品詞置き換えのみの手法が+11 種類で最良だった (表 6.4.19 参照)。

全手法において、出力結果から「うん」等使用頻度の高い応答が複数あった。例として提案手法の全テストデータ 694 発話に対する出力の、出力数上位 5 文を表 7.3.1 に示す。過去の発話を含まない学習の出力結果では、「うん」はテストデータの総数 694 のおよそ半分の 328 個である。過去の発話を含む学習データにおける出力では、同じ応答の使用頻度は全ての手法において減少した。

「うん」は文脈により意味、役割が変わりやすく、厳密な評価が困難である。また、表 7.3.1 の出力は全て、文脈により評価がわかれやすい応答である。非タスク指向型対話システムにおいて、似たような出力が多くなることは好ましくないと思われる。

表 7.3.1: 提案手法の出力数上位 5 文

過去の発話を含まない		過去の発話を含む	
出力 (応答)	出力数	出力 (応答)	出力数
うん	328	うん	258
ふーん	27	ふーん	31
うーん	16	うーん	22
そうですね	11	何?	13
はい	9	そうですね	11

第8章 おわりに

低頻度語処理によるニューラル対話モデルの性能向上のための研究を行った。

品詞に基づくヌルトークンの細分化と、評価のしやすさの向上のためのデータ調整を行った。ヌルトークンに品詞情報を付加することで細分化し、Copyable Model と組み合わせることで性能向上を試みた。また、過去の発話を学習・評価データとして用いることで評価のしやすさと厳密性の向上を試みた。

結果として、入力に過去の発話を含む学習データにおける精度評価の 評価は提案手法が最良で 0.34(68/200) となった。また、精度 評価の事例内における多様性評価は品詞付加のみの手法 (37 種類) と提案手法 (35 種類) が最良だった。精度 + 評価は品詞付加のみの手法と提案手法が最良であり、品詞付加のみの手法が 0.750(150/200) で提案手法が 0.755(151/200) となった。精度 + 評価内での多様性が最良の手法は低頻度語処理をしない手法 (75 種類) だった。また、過去の発話を付加することで評価のしやすさと厳密性が向上した。評価 200 文中、84 文の文脈が明確化した。

今後は、対話データの大幅な追加を行いつつ、出力の多様性と精度を向上させる。また、「うん」など多数出力される応答についての評価を厳格化する。具体的には、「うん」など出力数の多い応答の評価を 評価に固定するなど、多様性のあるモデルを高く評価できるように基準を見直す。

第9章 謝辞

最後に，一年間に渡り，本研究のご指導をいただきました鳥取大学工学部知能情報工学科自然言語処理研究室の村田真樹教授，村上仁一准教授に深く感謝すると共に，厚く御礼申し上げます．そして，日常の議論を通じて多くの知識や示唆を頂いた同研究室の皆様に深謝いたします．また，参考にさせていただいた論文の著者の方々に対して，深く感謝申し上げます．

参考文献

- [1] 佐藤 翔悦, 吉永 直樹, 豊田 正史, 喜連川 優 : 暗黙の発話状況を考慮したニューラル対話モデル, 言語処理学会第 23 回年次大会発表論文集, pp.374-377, 2017.
- [2] Ilya Sutskever, Quoc V.Le, Oriol Vinyals, Wojciech Zaremba : Addressing the Rare Word Problem in Neural Machine Translation, ACL 2015, pp.11-19, 2015.
- [3] 関沢 祐樹, 梶原 智之, 小町 守 : 目的言語の低頻度語の高頻度語への言い換えによるニューラル機械翻訳の改善, 言語処理学会第 23 回年次大会発表論文集, pp.982-985, 2017.
- [4] 村田 真樹, 内山 将夫, 白土 保, 井佐原 均 : シリーズ型質問文に対して単純結合法を利用した遞減的加点質問応答システム, システム制御情報学会論文誌, Vol.20, No.8, pp.338-346, 2007.
- [5] Ilya Sutskever, Oriol Vinyals, and Quoc V.Le. : Sequence to sequence learning with neural networks., Advances in Neural Information Processing Systems 27(NIPS2014), pp.1-9, 2014.
- [6] 対話破綻検出チャレンジ, <https://sites.google.com/site/dialoguebreakdownndetection/>
- [7] 名大会話コーパス, <http://mmsrv.ninjal.ac.jp/nucc/>
- [8] Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, Alexander M. Rush, “OpenNMT: Open-Source Toolkit for Neural Machine Translation”, ArXiv e-prints 1701.02810, 2017
- [9] Mecab : mecab-0.97.tar.gz , mecab-ipadic-2.7.0-20070801.tar.gz
<http://mecab.sourceforge.net/>.