特集論文 「知的対話システム」

キーワード・文脈情報の使い分けを可能にするデータ拡張を利用した未知語に対応する発話理解手法

Slot Filling with Data Augmentation That Allows the Use of Keyword and Context Information for Handling Unknown Slot Values

小林 優佳 Yuka Kobayashi
株式会社東芝 研究開発センター
Corporate R&D Center, Toshiba Corporation
yuka3.kobayashi@toshiba.co.jp

久島 務嗣 Tsuyoshi Kushima
（同 上）
tsuyoshi2.kushima@toshiba.co.jp

吉田 尚水 Takami Yoshida
（同 上）
takami.yoshida@toshiba.co.jp

藤村 浩司 Hiroshi Fujimura
（同 上）
hiroshi4.fujimura@toshiba.co.jp

岩田 憲治 Kenji Iwata
（同 上）
kenji4.iwata@toshiba.co.jp

keywords: spoken dialogue system, spoken language understanding, slot filling, unknown slot value

Summary

This paper proposes a new method for slot filling of unknown slot values (i.e., those are not included in the training data) in spoken dialogue systems. Slot filling detects slot values from user utterances and handles named entities such as product and restaurant names. In the real world, there is a steady stream of new named entities and it would be infeasible to add all of them as training data. Accordingly, it is inevitable that users will input utterances with unknown slot values and spoken dialogue systems must correctly estimate them. We provide a value detector that detects keywords representing slot values ignoring slots and a slot estimator that estimates slots for detected keywords. Context information can be an important clue for estimating slot values because the values in a given slot tend to appear in similar contexts. The value detector is trained with positive samples, which have keywords corresponding to slot values replaced with random words, thereby enabling the use of context information. However, any approach that can detect unknown slot values may produce false alarms because the features of unknown slot values are unseen and it is difficult to distinguish keywords of unknown slot values from non-keywords, which do not correspond to slot values. Therefore, we introduce a negative sample method that replaces keywords with non-keywords randomly, which allows the slot estimator to learn to reject non-keywords. Experimental results show that the proposed method achieves an 6,15 and 78% relative improvement in F1 score compared with an existing model on three datasets, respectively.

1. は じ め に

近年，携帯端末や AI スピーカー上で音声によるやりとりでユーザのアシストをする音声対話システムが注目されている．従来のボタン操作の UI とは異なり，ユーザは自然で自由な発話でシステムとやりとりすることで目的を達成することができる．本稿ではタスク指向対話システムに着目する．タスク指向対話システムは店舗検索，カーナビゲーション，家電機器操作などの特定のドメインにおいて，ユーザの要望を実現するために対話を行うシステムである．近年，様々なタスク指向音声対話システムが提案されている [Dahl 94, Kayama 10, Nisimura 05, Raux 05, Young 13].

ユーザが自然な発話でシステムに指示できるようにするためには，ユーザ発話から必要な情報を正確に抽出する必要がある．本稿ではユーザ発話からスロット・バリューを抽出するスロットフィリング発話理解手法に着目する．バリューとは対話システムが処理を実行するうえで必要な語句で，スロットはバリューの属性を表す．例えば，レストラン検索ドメインにおいて「安い和食が食べたい」

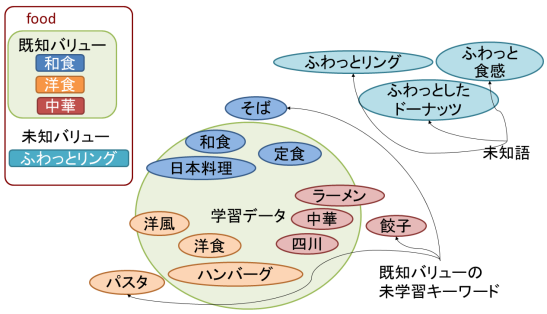

図 1 未知バリュー・未知語の例

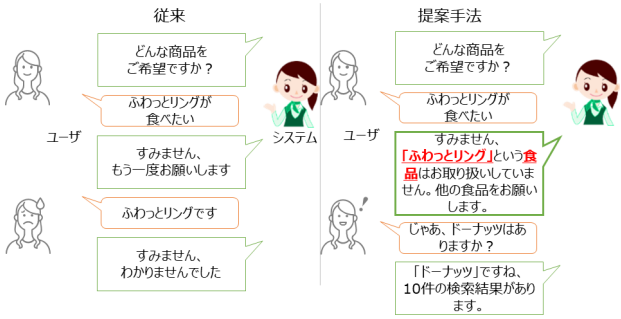

図 2 未知語の対話例

という発話からは (スロット：food，バリュー：和食)，(スロット：pricerange，バリュー：安い) というスロット・バリューが抽出される．

発話理解手法は，ルールベース [Jackson 91, Seneff 92] や統計ベース [He 03, McCallum 00, Raymond 07] の手法が提案されている．従来はルールベース手法が使用されていたが，開発コストがかかる，ルール作成に専門知識が必要という問題点があり，統計ベース手法が注目されている．

統計ベース手法で推定するためには，対象のスロット・バリューを扱ったユーザ発話を収集する必要がある．収集時に存在しなかったスロット・バリューは学習データに含まれない．本稿では，学習データ収集時に対象となったスロット・バリューを既知バリュー，それ以外のスロット・バリューを未知バリューと呼ぶ．図 1 では「洋食」「和食」「中華」というバリューを収集対象とした例であるが，この 3 つとは異なる「ふわっとリング」という未知バリューが存在している．

また，ユーザはバリューを表現するために様々な言い回しを使用する．本稿ではバリューを表現する言い回しをキーワードと呼ぶ．すべての言い回しを収集することは困難なので，図 1 のように学習データに含まれないキーワードが存在する．また，未知バリューのキーワードは収集していないので学習データに含まれない．本稿では，既知バリューのキーワードで学習データに含まれないもの (図では「パスタ」「そば」「餃子」) を未学習キーワード，未知バリューのキーワード (図では「ふわっとしたドーナッツ」「ふわっと食感」など) を未知語と呼ぶ．未学習キーワードは学習データ内のキーワードと類似していることが多い．一方で，未知語は収集対象ではないバリューのキーワードなので，学習データ内のキーワードとは類似していないことが多く，未学習キーワードよりも推定が困難である．

対話システムが未知語を理解できないと，図 2 左のようにユーザ発話に含まれる未知語を無視してしまう．ユーザはなぜ伝わらないのかわからず，対話が進まない．図 2 右のように未知語を「知らない」ということをユーザに正しく伝えることができれば，ユーザは別のキーワードでシステムに伝えることができるようになり，対話を円滑に進めることができる．またシステムにとっても「ふわっとリング」が学習データに含まれていないということがわかれば，学習データに追加することができ，効率よく未知語を減らすことができる．そのため未知語を扱える発話理解手法は非常に重要である．

近年，様々な言い回しを理解するために word2vec [Mikolov 13] や BERT [Devlin 19] などの単語分散表現が利用されている．単語分散表現を使うと，学習データ内のキーワードと類似している未学習キーワード・未知語を理解することができる．しかし，バリューには商品名・店舗名などの固有名詞が含まれ，これらはもともとの単語の意味と異なる使われ方をすることもあり，学習データに含まれない固有名詞は単語分散表現を用いても理解できないことがある．また，スロット・バリューになりうる商品名，流行語などを逐次収集し，再学習するのは非常にコストがかかる．

また，発話理解はユーザ音声が音声認識によってテキストに変換され，形態素解析によって単語分割が行われた後の処理になる．音声認識・形態素解析・単語分散表現は辞書を用いて処理を行っており，辞書に含まれない単語は正しく解析することができない．発話理解手法はこれらの誤りを含んだ入力を正しく解析する必要がある．本稿では，テキスト入力された文や音声の書き起こし文を対象とすることで，音声認識誤りの影響は除外し，形態素解析や単語分散表現の誤りを含む入力について検討する．

本稿では既知スロットの未知バリューを表現するキーワード (未知語) を扱うことができる発話理解手法を提案する．発話文中のキーワード以外の部分を文脈と呼び，文脈の特徴量を使用して未知語を検出する．同じスロットのバリューは同じ文脈で使用されることが多い．これは未知バリューについても同様である．例えば食べ物名であれば，既知バリューの発話文「和食が食べたい」「和食の食べられるお店を教えて」の文脈は未知バリューに対しても「ふわっとリングが食べたい」「ふわっとリングの食べられるお店を教えて」のようにそのまま使うことができる．ただし，文脈情報を重視して検出すると，同じ文

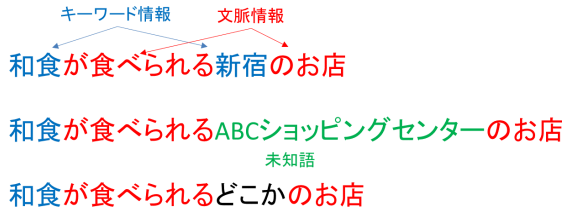

図 3 文脈情報を使用した推定例

脈であればキーワード部分にどんな単語がきても検出してしまい，図 3 の 3 つ目の例のように「どこか」を誤検出する可能性がある．そのため，誤検出した非キーワードをリジェクトする必要がある．そこで，文脈情報を利用して未知語を検出しつつ，誤検出を防ぐ方法を提案する．

発話理解タスクは発話からバリューを表現するキーワードを検出するタスクとスロットを推定するタスクに分解することができる．Liu らはバリュー検出器・スロット推定器を別々にモデル化・学習し，パイプラインでつないだ手法を提案している [Liu 20]．本稿でも同様に 2 つのタスクのモデルを個々に学習する．

バリュー検出器が文脈を重視して検出できるようにするために，学習データを拡張する [Kobayashi 18]．学習データ中のキーワードをランダムに別の単語に置き換えた文を生成し，学習データに追加する．こうすることで，学習データ中のキーワードには共通の特徴がなくなるため，学習されたモデルはキーワードではなく文脈の特徴量を重視したモデルとなる．本稿では追加された文をポジティブサンプルと呼ぶ．

前述のように，バリュー検出器は非キーワードを誤検出する場合がある．そこで，後段のスロット推定器で誤検出をリジェクトする．非キーワードは図 3 の「どこか」のように文脈の一部が誤検出されたものである．本稿では文脈中で使われる頻度の高い単語を文脈語と呼ぶ．バリュー検出器が検出した箇所の単語がすべて文脈語だった場合，誤検出の可能性が高い．そこで，キーワード中の各単語が文脈語かどうかの確率を推定する文脈語推定器を用意し，その中間特徴量をスロット推定器で使用する．文脈は，バリューが既知であるか未知であるかにかかわらず共通で使われる場合が多いので，文脈語であるかどうかを使用することでキーワードの箇所と誤検出された箇所を区別することができる．

また，スロット推定器がリジェクトを学習するための学習データを追加する．ポジティブサンプルではキーワードをランダムな単語に置き換えたが，ここでは文脈語で置き換える．置き換えた部分はキーワードではなく文脈として扱う．この文をネガティブサンプルと呼ぶ．ネガティブサンプルはオリジナルの文と文脈は同じでキーワード部分だけが異なるので，同じ文脈でもキーワードによってリジェクトする必要があることをスロット推定器が学習できる．

また，スロットは文脈情報だけでは特定できない場合がある．例えば「和食がいい」（スロット：food）「新宿がいい」（スロット：area）「子供向けがいい」（スロット：user）は異なるスロットだが，文脈は全く同じである．そこで，スロット推定器はバリュー検出器によって分割されたキーワードと文脈の特徴量を重みづけ平均して使用することで，キーワードと文脈のどちらを重視するかを発話文によって使い分けできるようにする．

またスロットには特定のキーワードのみではバリューが表現できないものもある．例えば DSTC3 データセット [Henderson 14] には true/false/don't care の 3 種類のバリューを持つ childrenallowed というスロットがあり，3 つのバリューはそれぞれ
“No I want a cheap pub with no children.”
“I am looking for a pub that allows children.”
“I do not care whether or not they allow children.”
といった発話文で表現される．こういったスロットは，文全体でバリューを判断する必要があるため，クラス分類タスクとして解く手法を適用し，発話文からスロット・バリューを抽出する手法と組み合わせる手法が提案されている [Gao 19]．また，上記のような true/false で表現されるスロットは未知バリューの出現の可能性が低い．そのため，Gao らの手法のようにクラス分類タスクの手法と組み合わせて推定することを想定し，本稿では上記のようなスロットは対象外とする．

本稿では提案手法について英語・日本語のコーパスで効果を確認した．本稿では，2 章で未知語を検出するための関連研究について述べる．3,4 章で提案手法について述べ，5 章で実験条件について述べ，6 章で実験結果について述べる．最後に 7 章で結論を述べる．

2. 関 連 研 究

時系列データの各要素にラベルをつける手法として Conditional Random Field (CRF) [Lafferty 01] が提案されている．また，機械翻訳で使用されている Recurrent Neural Network (RNN) エンコーダ・デコーダモデル [Cho 14, Sutskever 14] を用いた手法や，注意機構 [Bahdanau 16] を用いた手法 [Chen 16, Simonnet 15, Zhang 17, Zhao 18, Zhong 18] が提案されている．

1 章で述べたように，単語分散表現は既知の単語の言い換えなどには強いが，辞書に含まれていない単語は正しく解析できない．その弱点を補うための手法として，文字分散表現 [Bharadwaj 16, Chiu 16, Jaech 16, Makazhanov 16] や，サブワード [Abujabal 19, Egorova 18, Ribeiro 18] が提案されている．これらの分散表現の手法は発話理解の学習データ内の単語と近い未知語を正しく解析することができる．しかし，商品名などの固有名詞は本来の単語の意味とは全く異なる表現の場合があり，単語分散表

現だけでは解析が難しい．

同じスロットのバリューは同じ文脈で扱われることが多い．この傾向を利用して未知語を検出する手法が提案されている．Delexicalization [Luong 15, Rajendran 18, Shin 18, Wang 18] はあらかじめキーワードの表を用意しておき，発話文中のキーワードをスロットごとに特定のシンボルに変換し，シンボルの種類と文脈を使用して学習する手法である．発話理解の学習データに入っていない単語でも，キーワードの表に入っていれば扱うことができる．しかし，キーワードの表に入っていない単語は扱えない．

また，キーワードと文脈の情報を併用して，未知語を解析する手法が提案されている [Ishiwatari 19, Zhao 18].

バリューの箇所を特定してからスロットを推定する Liu らの手法はドメイン転用の手法として多用されている [Hou 20, Liu 21, Siddique 21]．また，発話文にスロット・バリューが含まれているかどうかを判断するモデル，その箇所を特定するモデルを組み合わせる手法が提案されている [Wang 20, Wu 19].

また，発話文中のキーワードを置き換えてデータを拡張する手法が提案されている．キーワードと類似した単語に置き換えて，モデルが理解できるキーワードの範囲を拡張する手法 [Kolachina 17, Prange 15] や，バリューを同じスロットの他のバリューと置き換える手法 [Louvan 20] や，キーワードを学習データ内の別のキーワードと置き換えて，ネガティブサンプルとして学習に使用する手法 [Hou 18a] が提案されている．また，機械翻訳の手法を用いて学習データから新しい文を生成する手法 [Hou 18b] や，敵対的サンプルを用いてスペルミスや音声認識誤りを模擬したデータを生成して学習データに追加する手法も提案されている [Cao 20]．また，学習時にネガティブサンプルを使用する手法も提案されている [He 21, Jiang 21, Lee 21].

3. 対話システム

図 4 にシステム構成を示す．ユーザ発話が入力されると，まずバリュー検出器が発話からバリューを表現するキーワードを検出する．次に，検出したキーワードのスロットをスロット推定器が推定する．検出されたスロット・キーワードをシステムが保持しているオントロジー（スロット・バリューのリスト）と比較し，既知かどうかを判定する．既知であれば対応するバリューを特定する．対話制御部では既知であればコンテンツデータベースを検索し，検索結果に応じた応答を決定し，応答生成部でシステム応答文を返す．未知であれば「すみません,「ふわっとリング」という食品はお取り扱いしていません．他の食品をお願いします.」のような応答文で対応できないことをユーザへ知らせる．本稿ではバリュー検出・スロット推定の部分について述べる．既知かどうかの判定は単

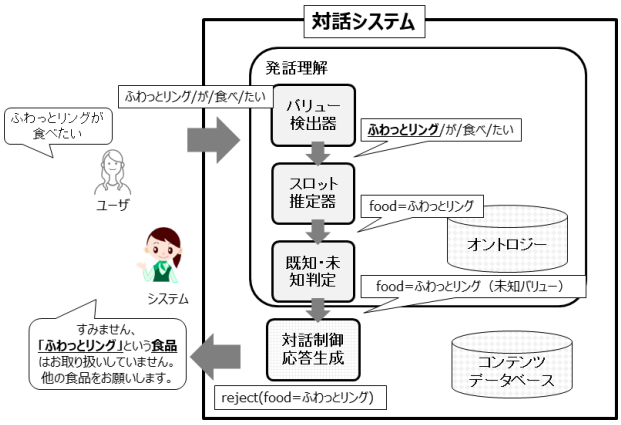

図 4 対話システム構成

語分散表現などを用いて，検出されたスロット・キーワードとオントロジー内のスロット・バリューの類似度を算出し，あらかじめ決めておいた閾値よりも類似度が高ければ既知と判定する．同様に類似度を元にバリューを特定する．

4. 提案手法

スロットフィリングタスクには IOB ラベルを用いることができる．図 5 は IOB ラベルの例である．キーワード以外の単語には「O」，キーワードの 1 単語目には「B-SLOT」(SLOT はスロット名)，2 単語目以降には「I-SLOT」を付与する．ユーザ発話の単語列を入力して，IOB ラベル列を出力する系列ラベリングタスクとして解くことができる．また，各スロットのキーワードの開始・終了位置 (スパン) を推定するスパン予測タスクとして解くこともできる．

本研究では，IOB ラベルを用いる．まず，バリュー検出器でスロットを無視した IOB ラベルを系列ラベリングタスクとして解く．次に検出されたキーワードのスロットを分類タスクとして解く．IOB ラベルを用いる理由を以下に述べる．スパン予測タスクを用いた手法には，1 発話に各スロットのスパンが 1 組あることを前提としてスパンの最適解を推定する手法 [Heck 20, Ouchi 18, Wu 19] がある．バリュー検出器はスロット名を無視することですべてのスロットが 1 つにまとめられるため，1 つの発話文に複数のキーワードが含まれることが多い．そのため，複数のスパンを推定する複雑なタスクとして解くことが必要になる．それに対して IOB ラベルを用いると，各単語に対して最適なラベルを求めるタスクとして解くことが可能なので，本稿では IOB ラベルを用いて推定する．

4·1 パイプラインモデル

バリュー検出・スロット推定を別々のモデルで行い，パイプラインでつなぐ (図 6)．学習は次のように行われる．

⟨s⟩	チョコ	ケーキ	が	食べ	たい	⟨/s⟩
O	B-FOOD	I-FOOD	O	O	O	O

⟨s⟩	高級な	ケーキ	が	ほしい	⟨/s⟩
O	B-PRICERANGE	B-FOOD	O	O	O

図 5 IOB ラベル付与例

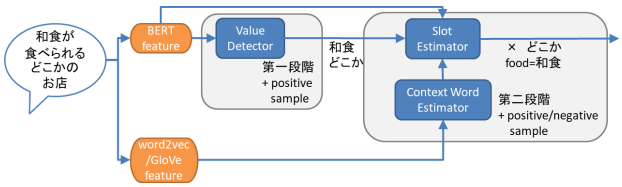

図 6 パイプラインモデル 構成

バリュー検出器の学習を行い (第一段階)，最も高性能なモデルを選択し，そのモデルの推定結果を用いてスロット推定器の学習を行う (第二段階)．スロット推定器の学習時はバリュー検出器のパラメータは固定する．

4·2 バリュー検出器

系列ラベリングは次式のように単語列 $x = (x_1, x_2, x_3, ..., x_N)$($N$:発話文中の単語数) から IOB ラベル列 $y = (y_1, y_2, y_3, ..., y_N)$ を推定するタスクである．

$$\hat{y} = \arg\max_{y} P(y|x) \tag{1}$$

バリュー検出器には RNN エンコーダ･デコーダベースのモデル [Liu 16] を用いる (図 7)．バリュー検出器はスロットを区別する必要がないので，出力する IOB ラベルはスロット名を含まない I,O,B の 3 種類である．エンコーダは発話文全体を考慮できるように双方向 LSTM(BDLSTM) [Schuster 97] を使用する．デコーダには LSTM を使用する．エンコーダは順方向 LSTM と逆方向 LSTM を結合して，単語列 x から隠れ状態 $h^d = (h_1^d, h_2^d, h_3^d, ..., h_N^d)$ を生成する．

$$f_i^d = E(x_i) \tag{2}$$

$$h_i^d = \mathrm{BDLSTM}(f_i^d, h_{i-1}^d, h_{i+1}^d) \tag{3}$$

単語分散表現 E には BERT を使用する．エンコーダの最終状態 h_N^d をデコーダの初期状態として使用する．エンコーダの隠れ状態から注意機構の重みを算出しコンテキストベクトル c_i を算出する．デコーダの隠れ状態 s_i^d は一つ前の隠れ状態 s_{i-1}^d，一つ前の出力 y_{i-1}^d，コンテキストベクトル c_i，エンコーダの隠れ状態 h_i^d から算出される．

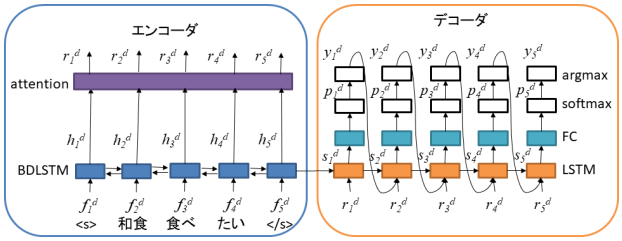

図 7 バリュー検出器 ネットワーク構成

表 1 ポジティブ・ネガティブサンプル例

オリジナル	**チョコ**	**ケーキ**	が	食べ	たい
	B-FOOD	I-FOOD	O	O	O
ポジティブ	**パソコン**	**空**	が	食べ	たい
サンプル	**夏**	**歌**	が	食べ	たい
	夕暮れ	**USB**	が	食べ	たい
	O	O	O	O	O
ネガティブ	**が**	**いい**	が	食べ	たい
サンプル	**教え**	**どこか**	が	食べ	たい
	探し	**いい**	が	食べ	たい

$$g_{i,j} = W_a(h_i^d \oplus h_j^d) + b_a \tag{4}$$

$$\alpha_i = \mathrm{softmax}(g_i) \tag{5}$$

$$c_i = \sum_{j=1}^{N} \alpha_{i,j} h_j^d \tag{6}$$

$$r_i^d = W_b(h_i^d \oplus c_i) + b_b \tag{7}$$

$$s_i^d = \mathrm{LSTM}(r_i^d \oplus y_{i-1}^d, s_{i-1}^d) \tag{8}$$

$W_{\{a,b\}}$ は重み，$b_{\{a,b\}}$ はバイアスを表す．$\oplus$ はベクトルの結合を表す．また，softmax() の式は下記で表される．

$$\mathrm{softmax}(x) = [\hat{x}_1, \hat{x}_2, \hat{x}_3, \ldots \hat{x}_N] \tag{9}$$

$$\hat{x}_i = \frac{\exp(x_i)}{\sum_{k=1}^{N} \exp(x_k)} \tag{10}$$

デコーダの LSTM の隠れ状態 s_i^d を全結合層 (FC) に入力し，その出力を softmax 層に入力し，各ラベルに対する確率 p_i^d が得られる．

$$p_i^d = \mathrm{softmax}(W^d s_i^d + b^d) \tag{11}$$

W^d は重み，b^d はバイアスを表す．バリュー検出器はスロット名を含まない IOB ラベルを推定するため，確率 p_i^d は $p_i^d = \left[p_{i,B}^d, p_{i,I}^d, p_{i,O}^d\right]$ となる．確率 $p_{i,j}^d$ から各単語に対するバリュー検出器の出力ラベル y_i^d が算出される．

$$y_i^d = \arg\max_{j \in \{B,I,O\}} p_{i,j}^d \tag{12}$$

4·3 ポジティブサンプル

1 章で述べたように未知バリューであっても，既存スロットのバリューであればユーザは同じ文脈で使用することが多いので，文脈特徴量を使用すると未知語も検出することができる．そこで，文脈特徴量を使用して検出するようにモデルを学習させる．本手法では学習データ中で B-SLOT,I-SLOT ラベルがふられている単語をラン

ダムに別の単語に置き換える (表 1)．こうすることで，学習データ中のキーワードには共通の特徴がなくなるため，モデルはキーワード特徴量に依存しないモデルとなる．モデルはキーワード以外の部分，すなわち発話文の文脈特徴量を重視したモデルとなる [Kobayashi 18].

ただし，任意の単語でキーワード部分を置き換えると，文脈中で使用されている単語で置換されて「食べたいが食べたい」のようにどちらの「食べたい」が文脈なのかわからない文が生成される可能性がある．そこであらかじめ学習データの文脈中で使用されている単語 (文脈語) をリストアップしておき，文脈語以外の単語で置換するようにしておく．具体的には学習データ中の各単語について O ラベルが付与されている頻度を算出しておき，全体の単語数に対する割合を算出し，割合が閾値 (0.00001) 以上であれば文脈語として扱う．単語分散表現の辞書中の単語で，文脈語ではないものを任意に選択して文を生成する．この手法で生成した発話文をポジティブサンプルと呼ぶ.

BERT は文全体をエンコードするので，キーワード部分だけを置換しても，ポジティブサンプルの文脈部分の特徴量はオリジナル文の特徴量と異なるものになる．ポジティブサンプルの文脈部分の特徴量をオリジナル文のものと近づけるために下記のロスを加える．オリジナル文の単語列を $x = (x_1, x_2, x_3, ..., x_N)$, この文から生成されたポジティブサンプルの単語列を $\hat{x} = (\hat{x}_1, \hat{x}_2, \hat{x}_3, ..., \hat{x}_N)$ とする．BERT のトークナイザは文の最初に [CLS] トークンを配置し，最後に [SEP] トークンを配置する．オリジナル文とポジティブサンプルの [CLS] トークンに相当する位置の BDLSTM の隠れ状態が近くなるように下記のロス L_m を y_i^d のクロスエントロピーロスに加える．

$$L_m = |h_0^d - \hat{h}_0^d| \tag{13}$$

4·4 スロット推定器

スロット推定器はバリュー検出器によって検出されたキーワードのスロットを推定する．キーワードの箇所は特定できているので，該当キーワードのスロットを分類する分類問題として解く．1 つの発話文に複数のキーワードが含まれている場合はキーワードの数だけ発話文を複製して，各キーワードのスロットを推定する．

§1 文脈語推定器

バリュー検出器は文脈情報を使用してキーワードを検出するため，キーワード部分にどんな単語が入っていても検出する．未知語が検出できるようになる一方で，誤検出が増える．検出された箇所がすべて文脈で使われる単語である場合は誤検出である可能性が高い．文脈で使われているかどうかは 4·3 節で使用した文脈語を使用して表現する．各単語について文脈語かどうかのラベル y_i^c を推定する文脈語推定器を用意し，文脈語推定器の中間

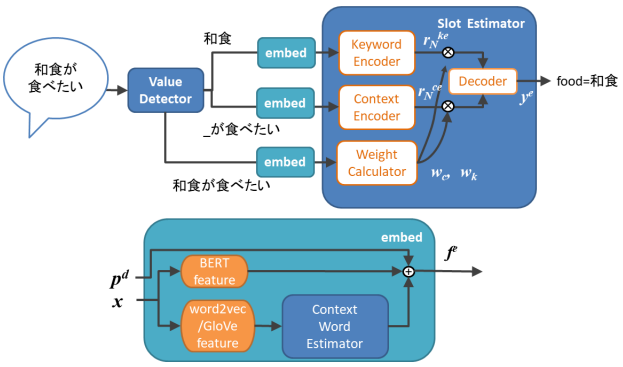

図 8 スロット推定器 ネットワーク構成

特徴量を使用する．y_i^c は下記で定義される．

$$y_i^c = \begin{cases} 1 & (\text{if } x_i \text{ is a context word}) \\ 0 & (\text{or not}) \end{cases} \tag{14}$$

文脈語推定器は y_i^c を下記のように推定する．

$$h_i^c = W_2^c(W_1^c E_w(x_i) + b_1^c) + b_2^c \tag{15}$$

$$p_i^c = \frac{1}{1 + \exp(W_3^c h_i^c + b_3^c)} \tag{16}$$

$$y_i^c = \begin{cases} 1 & (\text{if } p_i^c > 0.5) \\ 0 & (\text{or not}) \end{cases} \tag{17}$$

モデルの入力には単語 x_i を単語分散表現に変換した $E_w(x_i)$ を使用する．$W_{\{1,2,3\}}^c$ は重み，$b_{\{1,2,3\}}^c$ はバイアスである．文脈語推定器は各単語について文脈語かどうかを推定するので，単語の特徴量のみを使用し，前後の文脈の特徴量は使用しない．そこで，E_w は文全体でエンコードする BERT ではなく，単語単体でエンコードする word2vec を使用する．

§2 スロット推定器 ネットワーク

バリュー検出器によってキーワードの箇所がわかっているので，発話文をキーワード部分とそれ以外の文脈部分に分けてエンコードする (図 8)[Kobayashi 19]．2 つのエンコーダの最終状態を重み加算してデコードしてスロットを推定する．スロット推定器の入力 f_i^e は文を BERT でエンコードした特徴量，バリュー検出器の出力確率 p_i^d, 文脈語推定器の中間特徴量 h_i^c を結合したものになる．

$$f_i^e = embed(x_i, p_i^d) \tag{18}$$

$$= E(x_i) \oplus p_i^d \oplus h_i^c \tag{19}$$

キーワード情報 f_i^{ke} はバリュー検出器によってキーワード箇所と特定された箇所だけ f_i^e を残し，それ以外の部分は同じ次元のゼロベクトルにする．反対に文脈情報 f_i^{ce} はバリュー検出器によって文脈と特定された箇所だけ f_i^e を残し，それ以外の部分は同じ次元のゼロベクトルにする．

$$h_i^{ke} = \text{BDLSTM}(f_i^{ke}, h_{i-1}^{ke}, h_{i+1}^{ke}) \tag{20}$$

$$h_i^{ce} = \text{BDLSTM}(f_i^{ce}, h_{i-1}^{ce}, h_{i+1}^{ce}) \tag{21}$$

$$r_i^{ke} = attention(h_i^{ke}, h^{ke}) \tag{22}$$

$$r_i^{ce} = attention(h_i^{ce}, h^{ce}) \tag{23}$$

$attention$ は式 (4)-(7) をまとめたものである．次に発話文の情報を使用して，各エンコーダの出力に付与する重み w_c, w_k を算出し，r_N^{ke}, r_N^{ce} を重みづけ加算する．

$$h_i^e = \mathrm{BDLSTM}(f_i^e, h_{i-1}^e, h_{i+1}^e) \quad (24)$$

$$r_i^e = attention(h_i^e, h^e) \quad (25)$$

$$[w_c, w_k] = W^f r_N^e + b^f \quad (26)$$

$$q = w_c r_N^{ce} + w_k r_N^{ke} \quad (27)$$

$$p^e = \mathrm{softmax}(W^g q + b^g) \quad (28)$$

$$y^e = \arg\max_j p_j^e \quad (29)$$

$W^{\{f,g\}}$ は重み，$b^{\{f,g\}}$ はバイアスである．p^e はスロットの数 N_{slot} 次元のベクトルになり，各要素は各スロットに対する確率になる．

4·5 ネガティブサンプル

学習時にスロット推定器がキーワードをリジェクトするべきデータがなければ，スロット推定器はリジェクトする挙動を学習できない．そこで，リジェクトする必要があるデータを用意する．ポジティブサンプルを生成したときと同様に，学習データ中のキーワードを文脈語群からランダムに選択した単語に置き換える．置き換えられた部分の IOB ラベルは O にする (表 1)．生成した発話文をネガティブサンプルと呼ぶ．ネガティブサンプルと元の文は同じ文脈を持つので，同じ文脈でもキーワード部分が文脈語の場合はリジェクトするようにスロット推定器が学習される．

5. 評 価 実 験

5·1 デ ー タ セット

ここでは，Simulated Dialogue[Shah 18](英語) と独自に収集したショッピングセンター店舗検索 (日本語)(以降 Shop) のデータセットを使用する．Simulated Dialogue は映画のチケットを購入する対話 (Sim-M)，レストランを予約する対話 (Sim-R) の 2 種類のデータセットがある．学習・検証・テストデータが用意されている．それぞれ movie (映画名) スロット，restaurant_name (レストラン名) スロットに未知バリューが含まれている．

Shop はクラウドソーシングで収集したショッピングセンター内の店舗を検索する対話コーパスで表 2 の 5 種類のスロットがある．最初に，検索する店舗の条件を提示し，ルールベースで作成した対話システムと対話する形式で発話文をテキスト入力するタスクを実施し，対話データを収集した (タスク 1)．また，未知バリューを収集するために item,food,name スロットについてバリューとそのバリューで検索するための発話文を入力するタスクを実施した (タスク 2)．バリューはユーザに自由に考えてもらい，実在する商品名や店舗名を含む発話文を収集した．タスク 1 で収集したデータを学習・検証・テストに分割し，さらにタスク 2 で収集したデータをテストデータに追加した．表 3 に収集した発話例を示す．

表 2 Shop コーパス 学習データスロット・バリュー例

スロット名	バリュー例
item	家具，家電，服，お菓子，お酒 ···
genre	おみやげ，レストラン，生活雑貨 ···
food	和食，パスタ，カフェ，スイーツ ···
user	男性，女性，子供 ···
name	○○コーヒー，△△食堂 ···

(name スロットは実際のデータでは実在する店舗名が使われている)

表 3 Shop コーパス 収集した発話例

(a) タスク 1

システム： ご利用ありがとうございます．ご用件をお願いします．
ユーザ： 女性用の服の店を探しています．
システム： 婦人向けの服のお店として「△△アウトレット」が見つかりました．
ユーザ： 場所はどこですか．
システム： 「△△アウトレット」の場所は南側エリア-11 です．
ユーザ： 電話番号は．
システム： 「△△アウトレット」の電話番号は 000-0000-0000 です．

(b) タスク 2

スロット名	発話例
food	**キッズラーメンセット**食べられるところ **ヨーグルトシェイク**が飲みたい **シーザーサラダ**のおいしい店
item	**ミニカー**を買いたい この近くで**ジグゾーパズル**が買える店 **イオンドライヤー**を買い直したい
name	○○**アウトレット**の行き方が知りたい ここから一番近い△△**ドラッグ**はどこ？ 豊中駅近くの□□**屋**に行きたい

未知語の発話理解には，学習データに含まれていないスロット・バリューを扱うという課題と，商品名・店舗名などの固有名詞は単語本来の意味と異なる使われ方をするために単語分散表現を用いても理解できないという課題がある．後者について検証するためには，固有名詞を含むコーパスが必要である．Sim-M,R は映画名，映画館名，レストラン名について実際の名前を模擬した名称が使用されており，同様に日本語で実在する商品名・店舗名を扱った評価を実施するために，Shop を独自に収集した．各データセットのスロット・バリューの数を表 4 に載せる．

表 4 データセットサイズ (括弧内はユニーク数)

(a) Sim-M

データ	発話文数	バリュー数	
		既知	未知
学習	1,973	2,059(76)	0
検証	627	523(47)	62(5)
テスト	1,364	1,187(51)	176(26)

(b) Sim-R

データ	発話文数	バリュー数	
		既知	未知
学習	6,175	5,415(65)	0
検証	1,489	1,140(63)	134(4)
テスト	3,436	2,763(63)	306 (11)

(c) Shop

データ	発話文数	バリュー数	
		既知	未知
学習	8,054	4,993 (104)	0
検証	1,303	842 (68)	5 (3)
テスト	8,053	2,538 (82)	4,359 (934)

5·2 評 価 手 法

バリュー検出性能とスロット推定性能をそれぞれ評価する．ここでは F 値を用いて評価する．F 値は下記で算出される．

$$\text{適合率} = \frac{\text{正解したキーワード数}}{\text{推定キーワード数}} \times 100 \quad (30)$$

$$\text{再現率} = \frac{\text{正解したキーワード数}}{\text{正解ラベルのキーワード数}} \times 100 \quad (31)$$

$$F\ \text{値} = \frac{2 \times \text{再現率} \times \text{適合率}}{\text{再現率} + \text{適合率}} \quad (32)$$

本稿では seqeval [*1] [Nakayama 18] を用いて算出した．

5·3 実 験 方 法

バリュー検出器・スロット推定器の単語分散表現 E には BERT を使用した．BERT の実装には PyTorch 1.7.0[*2], transformers 4.0.0[*3] [Wolf 20] を使用し，BERT モデルは transformers 内で用意されている `bert-base-uncased`(英語), `cl-tohoku/bert-base-japanese-v2`(日本語) を使用した．文脈語推定器の単語分散表現 E_w には英語モデルは GloVe [Pennington 14] を使用し，日本語モデルは独自に収集したウェブコーパスで学習した word2vec モデルを使用した．実装には gensim [*4] を使用した．

また，実験条件として Word dropout (WD)[Iyyer 15], Target feature dropout (TFD)[Xu 18] を使用する．また，注意機構のあり・なしも実験条件として使用する．注意機構を用いない場合はデコーダにはエンコーダの隠れ状態を入力する．

表 5 実験パラメータ

パラメータ	値
BERT 次元	768
GloVe 次元 (英語)	300
word2vec 次元 (日本語)	512
バッチサイズ	32
LSTM サイズ	128
隠れ層数	1
勾配クリッピング	1.0
Dropout の割合	0.3
学習率	0.001
荷重減衰	0.001
最適化手法	AdamW
Word dropout	0,0.3,0.5
Target feature dropout	0,0.3,0.5
ポジティブ・ネガティブサンプル	×1, ×3, ×5

ポジティブ・ネガティブサンプルは 1 つの発話文から複数の発話文を生成することが可能である．1 つの発話文から生成する文の数を倍率と呼び，倍率を変えて実験する．学習時にはオリジナルの発話文と生成された発話文を使用する．表 5 は学習時のパラメータである．Dropout は RNN 以外の層に適用する．TFD と WD は手法が似ているため，同時には適用しない．

各試行において乱数のシード値を 5 種類用意し，平均・分散で評価を行う．シードは，ポジティブ・ネガティブサンプルを生成する際とネットワークの初期値生成に用いる．また，WD,TFD を用いる際はドロップアウトする単語を選択する際に用いる．

6. 実 験 結 果

6·1 バ リュー 検 出

最初にバリュー検出に対する性能評価を行う．Sim-M,R には数字のキーワードのみを持つスロットがある (`num_people`, `time`, `num_tickets`)．これらのスロットは数字キーワードしか持たず，未知語を持つ可能性が低い．また，数字キーワードしか持たないという特徴を活用した方が推定しやすい．そこで，この 3 スロットについてはポジティブサンプルを生成しない．

表 6 は実験結果である．バリュー検出器はスロット名を無視して F 値を算出している．各実験条件において，乱数のシード値を 5 種類用意して試行し，5 回の平均値を各実験条件の値として用いる．実験条件の中で再現率の平均値が最も高かった実験条件について，平均値を載せている．バリュー検出で誤検出したとしてもスロット推定時にリジェクトすることができるが，検出漏れはスロット推定器では補うことができない．そのためバリュー

∗1 https://github.com/chakki-works/seqeval
∗2 https://pytorch.org/
∗3 https://huggingface.co/transformers/
∗4 https://radimrehurek.com/gensim/index.html

表 6 バリュー検出　実験結果 (VD: バリュー検出器, PS: ポジティブサンプル, ML: L_m 追加).

(a) Sim-M

モデル	適合率	再現率	F 値
VD	96.00	95.41	95.71
VD+PS	95.84	96.40	96.12
VD+PS+ML	**96.02**	**96.50**	**96.26**

(b) Sim-R

モデル	適合率	再現率	F 値
VD	**98.14**	93.03	95.52
VD+PS	95.13	**96.18**	**95.65**
VD+PS+ML	94.92	95.93	95.42

(c) Shop

モデル	適合率	再現率	F 値
VD	77.66	85.56	81.42
VD+PS	**79.33**	89.75	84.22
VD+PS+ML	79.10	**90.08**	**84.23**

検出器は高い再現率を持つことが重要になる．

実験結果を見ると，ポジティブサンプルを使用することで再現率が向上し，適合率が悪化したことがわかる．未知語が検出できるようになった一方で誤検出が増えているといえる．また L_m の効果を見ると Sim-M,Shop では効果が確認できたが，Sim-R では確認できなかった．

6·2 スロット推定

次にスロット推定の評価を行う．baseline として同じネットワーク (図 7) でスロット名を含む IOB ラベルを推定する手法を使用する．提案手法ではバリュー検出とスロット推定を別々のモデルで行うが，baseline は 1 つのモデルで行う．また，比較手法として Slot-Gated モデル [Goo 18] を使用する．

表 7 は実験結果である．ここではスロット名を考慮して F 値を算出している．F 値について baseline と t 検定を行い，$p < 0.05$ で有意差が認められた結果には*をつけている．テストデータ全体に対する性能とは別にテストデータ内の未知語の再現率を載せる．

スロット推定器学習時にポジティブ・ネガティブサンプルをそれぞれ適用，両方適用した試行を実施した．スロット推定器は VD+PS+ML のバリュー検出器モデルの中で未知語テストデータに対する再現率が最も高かった実験条件のモデル (表 6) を使用して学習し，実験結果の中で最も F 値が高かった結果を載せている．本来であればバリュー検出器のすべてのモデルでスロット推定器を学習し，実験結果の中で最も F 値が高いものを選択するべきであるが，性能が悪いバリュー検出器でスロット推定器を学習しても性能が出ないと思われるので，ここでは性能のよかった VD+PS+ML のモデルに限定して実験を行う．baseline, Slot-Gated も同様に最も F 値が高かっ

表 7 スロット推定　実験結果 (SE: スロット推定器, CWE: 文脈語推定器, NS: ネガティブサンプル, PNS:ポジティブ・ネガティブサンプル併用).

(a) Sim-M

	テスト全体			未知語
モデル	適合率	再現率	F 値	再現率
baseline [Liu 16]	**97.25**	95.64	96.44	78.26
Slot-Gated [Goo 18]	92.99	90.68	91.81	55.83
SE	96.02	**96.50**	96.26	**82.34**
SE + PNS	96.68	**96.50**	96.59	**82.34**
SE + CWE	96.02	**96.50**	96.26	**82.34**
SE + CWE + NS	96.76	96.44	96.60	82.04
SE + CWE + PS	96.40	**96.50**	96.45	**82.34**
SE + CWE + PNS	96.82	**96.50**	**96.66**	**82.34**

(b) Sim-R

	テスト全体			未知語
モデル	適合率	再現率	F 値	再現率
baseline [Liu 16]	**98.79**	92.76	95.68	30.53
Slot-Gated [Goo 18]	98.26	92.61	95.35	28.04
SE	95.76	95.28	95.52	56.93
SE + PNS	96.71	95.92	**96.32** *	**63.33**
SE + CWE	97.44	92.81	95.07 *	31.87
SE + CWE + NS	98.20	92.51	95.27 *	29.27
SE + CWE + PS	96.62	**95.93**	96.27 *	**63.33**
SE + CWE + PNS	96.71	95.92	96.31 *	**63.33**

(c) Shop

	テスト全体			未知語
モデル	適合率	再現率	F 値	再現率
baseline [Liu 16]	56.11	60.06	58.01	40.31
Slot-Gated [Goo 18]	60.13	37.40	45.91	9.57
SE	62.50	66.98	64.64 *	50.10
SE + PNS	90.82	89.97	90.39 *	90.08
SE + CWE	62.10	66.78	64.33 *	51.28
SE + CWE + NS	71.46	71.66	71.56 *	54.15
SE + CWE + PS	**91.33**	90.02	**90.67** *	**90.90**
SE + CWE + PNS	91.27	**90.04**	90.65 *	90.60

表 8 誤り分析

(a) Sim-M

モデル	スロット誤り	範囲誤り	誤検出
baseline	1	32	2
SE	0	47	11
SE + CWE + PNS	0	46	1

(b) Sim-R

モデル	スロット誤り	範囲誤り	誤検出
baseline	0	18	18
SE	11	89	10
SE + CWE + PNS	0	99	2

(c) Shop

モデル	スロット誤り	範囲誤り	誤検出
baseline	1657	826	703
SE	1590	635	567
SE + CWE + PNS	1	582	20

た結果を載せている．

また，誤り分析を行った．モデルが出力したキーワードの中で誤りとなったデータを 3 種類に分類する．

- スロット誤り：スロット推定の誤り
- 範囲誤り：正解キーワードと部分的に重複する部分の検出
- 誤検出：キーワードではない部分の検出

表 8 が各試行の誤り分析結果である．

baseline,Slot-Gated と比較すると，SE は性能が低い結果もあり，パイプラインにしただけでは効果が小さいが，SE+CWE+PNS はテスト全体 F 値，未知バリュー再現率ともに高性能であり，文脈語推定器とポジティブ・ネガティブサンプルを適用する提案手法が有効であることが確認できる．t 検定を実施した結果を見ると，Sim-M については有意差は認められなかったが，Sim-R と Shop については有意差を確認できた．

Sim-M,R について SE+CWE+NS と SE+CWE+PS を比較すると，NS の方が適合率が高く，PS の方が再現率が高い．ネガティブサンプルを適用することで誤検出したキーワードをリジェクトさせることができるようになった反面，未知語もリジェクトしてしまっているのが原因だと思われる．ポジティブサンプルはその逆で未知語をリジェクトしない反面，誤検出のリジェクトがあまりできない．SE+CWE+PNS はそれぞれのメリットをうまく融合させることができ，高性能になったのだと思われる．表 8 を見ると CWE,PNS を適用したことで誤検出が軽減できていることがわかる．

一方，範囲誤りは Sim-M,R では baseline よりも多い．baseline は学習データに含まれない未知語はほとんど検出しない．そのため，範囲が確実にわからないキーワードはほとんど検出しないので範囲誤りは少ない．それに対して提案手法では範囲が不確定な未知語も検出するため，範囲誤りが多く，その結果適合率が baseline より低くなっている．映画名やレストラン名は複数の単語からなる長い名称が多い．特に映画名は文章のような名称も含むため，範囲を特定することが難しい．提案手法で範囲誤りになったキーワードは baseline で正しく推定できているわけではなく，検出されていないことが多い．

Shop では SE+CWE+PS が最も高く，ポジティブサンプルを入れることで性能が向上していることがわかる．Sim-M,R は未知バリューは 1 つのスロットだけだったので，スロット推定時に未知バリューに特定のスロットを割り当てるだけでよかったが，Shop は未知バリューに適切なスロットを選択する必要がある．ネガティブサンプルはスロット推定の性能には寄与しないので，ネガティブサンプルだけではスロット誤りは軽減できず，ポジティブサンプルを適用することでスロット推定精度が向上したと考えられる．表 8 を見ても SE+CWE+PNS はスロット誤りが他より大幅に少なく，未知バリューのスロット推定性能が向上していることがわかる．

次に文脈語推定器の効果を確認する．SE+PNS と SE+CWE+PNS の F 値を比較すると，Sim-R では SE+PNS が高いが，他の 2 つのコーパスでは SE+CWE+PNS が高い．Sim-R では “good” というバリューがある．good は文脈中でも使用されるので，文脈語として扱われている．文脈語推定器を用いた推定では文脈語として扱われる単語はリジェクトするように学習するので，good はリジェクトされてしまうことが多く，やや性能が下がったと考えられる．それ以外のコーパスでは文脈語推定器をポジティブ・ネガティブサンプルと併用することで性能向上することができた．

次に，実験のパラメータの影響について考察する．表 A.1 に各コーパスで最も性能が高かった結果について，実験パラメータを載せる．TFD と提案手法の組み合わせはあまり効果がなかったことがわかる．提案手法はキーワード部分を任意の単語で置き換え，TFD はドロップアウトする．キーワード部分を任意に置き換えるという実装は似ているので，似たような手法を組み合わせても性能が上がらなかったと思われる．それに対して WD は文全体を任意にドロップアウトするので提案手法との組み合わせでより良い性能を出すことができた．また，表 A.2 にポジティブ・ネガティブサンプルの倍率を変えた結果を載せる．ポジティブ・ネガティブサンプルの倍率は高ければよいというわけではなく，最適な倍率があることがわかる．

次に，スロット推定器が正しくリジェクトできた例を示す．

- **Assistant**, please locate a restaurant for me.

- I need you **to look up** restaurants in middletown.
- **Perfect**.
- アウトレットで都心から**一番近いところ**を教えて.
- **今はやり**のタピオカミルクティーが飲めるところは.
- チーズケーキが食べられるお店で近い**とこ**は?

バリュー検出器が誤検出し，スロット推定器がリジェクトした部分を下線部分で示す.「洋食店を教えて」「近いカフェは？」などの文脈を学習したことによって上記が誤検出されたと考えられる．このように誤検出した部分をスロット推定器でリジェクトできることが確認できた.例文中の「とこ」「ところ」は代名詞で，参照先がキーワードの可能性がある．参照先が同じ発話内のキーワードであれば検出可能だが，以前の発話中にある場合は本手法では推定できない．そういった場合は対話履歴を用いて参照先を推定する必要がある.

6·3 形態素解析誤りの影響

冒頭で述べたように発話理解は音声認識・形態素解析処理の後に行われ，これらの誤りの影響を受ける．本実験はテキスト入力された文を扱っているので音声認識誤りは発生しないが，形態素解析誤りは発生する．また，単語分散表現のモデルにない単語が入力される可能性もある．本稿では日本語の BERT モデルに用いられている MeCab [*5]を用いて形態素解析を実施した．Shop コーパスについて形態素解析誤りがあるキーワードの数を目視で算出したところ，学習・検証データで 5 個，テストデータで 33 個あった．形態素解析はひらがな・カタカナが連続する単語の分割に失敗することが多く,「おこ/さま/ランチ」「タン/タンメン」などの誤りがあった．店舗名・商品名は本来漢字で表記される単語もひらがな・カタカナで表記される場合があり，形態素解析誤りが発生しやすい.形態素解析誤りがあると単語分散表現も誤った特徴量を付与する．この 33 個のキーワードを含む文の該当キーワードについて再現率を算出したところ，baseline では 31%,SE+CWE+PS では 65%だった．本手法では文脈に注目した解析を行っており，キーワード部分の形態素解析・単語分散表現の誤りの影響を受けにくいため，このように正しく推定することができた.

7. ま と め

未知語を扱うことができる発話理解手法として，キーワードと文脈を活用する手法を提案した．未知バリューであっても，既存スロットのバリューであれば同じ文脈で使用されることが多いため，文脈情報を使用することで未知語を検出することができる．しかし，文脈情報を重視して検出すると，同じ文脈であればキーワード部分にどんな単語がきても検出してしまい，誤検出が増えるという問題がある．本手法ではバリュー検出・スロット推定を別々のモデルで行い，パイプラインでつなぐ構成を使用し，学習データ中のキーワードをランダムに別の単語に置き換えるポジティブ・ネガティブサンプルを利用し，文脈情報を利用して未知語を検出し，誤検出を低減する手法を提案した.

Sim-M,Sim-R,Shop の 3 種類のデータセットで実験した結果，それぞれのデータセットで F 値が,96.66, 96.32, 90.67 となり，提案手法の効果を確認した．提案手法は再現率，特に未知語に対する再現率が高く，未知語が正しく検出できるようになったことがわかる．また，文脈に着目することで，キーワードの形態素解析誤り・単語分散表現の誤りの影響も低減できることを確認した.

一方で Sim-M,R において，適合率は baseline よりも低い結果となり，誤り分析の結果，範囲誤りが多いことがわかった．映画名やレストラン名などの長いキーワードの範囲を正しく推定することは今後の課題である.

今回は音声認識誤りは考慮せず実験を行ったが，今後は音声認識結果での性能評価も行っていく．音声認識のモデルに含まれていない未知語は正しく認識することができないので，発話理解モデルに入力されるキーワードが誤る可能性がある．キーワードが正しく音声認識できなかったとしても，前後の文脈が正しく認識できていれば推定することが可能である．ただし,「○○という食品はお取り扱いしていません.」という応答を返す際に「○○」は誤認識結果になるという問題点は残る．また，音声認識モデルに含まれていないから正しく音声認識できないのか，言い直せば正しく音声認識できるのかをユーザは判断することができない．今後はこういった誤認識を考慮した対話システムについても検討していく.

*5 https://taku910.github.io/mecab/

◇ 参 考 文 献 ◇

[Abujabal 19] Abujabal, A. and Gaspers, J.: Neural Named Entity Recognition from Subword Units, in *Proceedings of INTERSPEECH*, pp. 2663–2667 (2019)

[Bahdanau 16] Bahdanau, D., Cho, K., and Bengio, Y.: Neural Machine Translation by Jointly Learning to Align and Translate, in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 1054–1063 (2016)

[Bharadwaj 16] Bharadwaj, A., Mortensen, D. R., Dyer, C., and Carbonell, J. G.: Phonologically Aware Neural Model for Named Entity Recognition in Low Resource Transfer Settings, in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1462–1472 (2016)

[Cao 20] Cao, X., Xiong, D., Shi, C., Wang, C., Meng, Y., and Hu, C.: Balanced Joint Adversarialmi Training for Robust Intent Detection and Slot Filling, in *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 4926–4936, Barcelona, Spain (Online) (2020), International Committee on Computational Linguistics

[Chen 16] Chen, Y.-N., Hakkani-Tür, D. Z., Tür, G., Gao, J., and Deng, L.: End-to-End Memory Networks with Knowledge Carryover for Multi-Turn Spoken Language Understanding, in *Proceedings of INTERSPEECH* (2016)

[Chiu 16] Chiu, J. P. C. and Nichols, E.: Named Entity Recognition with Bidirectional LSTM-CNNs, *Transactions of the Association for*

Computational Linguistics, Vol. 4, pp. 357–370 (2016)

[Cho 14] Cho, K., Merriënboer, B., Çaglar, G., Bahdanau, D., Bougares, F., Holger, H., and Bengio, Y.: Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734 (2014)

[Dahl 94] Dahl, D. A., Bates, M., Brown, M., Fisher, W., Hunicke-Smith, K., Pallett, D., Pao, C., Rudnicky, A., and Shriberg, E.: Expanding the Scope of the ATIS Task: The ATIS-3 Corpus, in *Proceedings of Human Language Technology* (1994)

[Devlin 19] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, Minneapolis, Minnesota (2019), Association for Computational Linguistics

[Egorova 18] Egorova, E. and Burget, L.: Out-of-Vocabulary Word Recovery using FST-Based Subword Unit Clustering in a Hybrid ASR System, in *Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 5919–5923 (2018)

[Gao 19] Gao, S., Sethi, A., Agarwal, S., Chung, T., and Hakkani-Tur, D.: Dialog State Tracking: A Neural Reading Comprehension Approach, in *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue*, pp. 264–273, Stockholm, Sweden (2019), Association for Computational Linguistics

[Goo 18] Goo, C.-W., Gao, G., Hsu, Y.-K., Huo, C.-L., Chen, T.-C., Hsu, K.-W., and Chen, Y.-N.: Slot-Gated Modeling for Joint Slot Filling and Intent Prediction, in *Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (2018)

[He 03] He, Y. and Young, S.: A data-driven spoken language understanding system, in *Proceedings of 2003 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, pp. 583–588 (2003)

[He 21] He, H., Lu, H., Bao, S., Wang, F., Wu, H., Niu, Z., and Wang, H.: Learning to Select External Knowledge with Multi-Scale Negative Sampling, *arXiv preprint arXiv:2102.02096* (2021)

[Heck 20] Heck, M., Niekerk, van C., Lubis, N., Geishauser, C., Lin, H.-C., Moresi, M., and Gasic, M.: TripPy: A Triple Copy Strategy for Value Independent Neural Dialog State Tracking, in *Proceedings of the 21st Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 35–44, 1st virtual meeting (2020), Association for Computational Linguistics

[Henderson 14] Henderson, M., Thomson, B., and Williams, J.: The Third Dialog State Tracking Challenge, in *Proceedings IEEE Spoken Language Technology Workshop (SLT)*, pp. 324–329 (2014)

[Hou 18a] Hou, M., Wang, X., Yuan, C., Yang, G., Hu, S., and Shi, Y.: Attention Based Joint Model with Negative Sampling for New Slot Values Recognition, in *Proceedings of the 9th International Workshop on Spoken Dialogue Systems Technology (IWSDS 2018)*, pp. 3–15 (2018)

[Hou 18b] Hou, Y., Liu, Y., Che, W., and Liu, T.: Sequence-to-Sequence Data Augmentation for Dialogue Language Understanding, in *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 1234–1245, Santa Fe, New Mexico, USA (2018), Association for Computational Linguistics

[Hou 20] Hou, Y., Che, W., Lai, Y., Zhou, Z., Liu, Y., Liu, H., and Liu, T.: Few-shot Slot Tagging with Collapsed Dependency Transfer and Label-enhanced Task-adaptive Projection Network, in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 1381–1393, Online (2020), Association for Computational Linguistics

[Ishiwatari 19] Ishiwatari, S., Hayashi, H., Yoshinaga, N., Neubig, G., Sato, S., Toyoda, M., and Kitsuregawa, M.: Learning to Describe Unknown Phrases with Local and Global Contexts, in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT2019)*, pp. 3467–3476 (2019)

[Iyyer 15] Iyyer, M., Manjunatha, V., Boyd-Graber, J., and Daumé III, H.: Deep Unordered Composition Rivals Syntactic Methods for Text Classification, in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pp. 1681–1691 (2015)

[Jackson 91] Jackson, E., Appelt, D., Bear, J., Moore, R., and Podlozny, A.: A Template Matcher for Robust NL Interpretation, in *Proceedings of Speech and Natural Language*, pp. 190–194 (1991)

[Jaech 16] Jaech, A., Heck, L., and Ostendorf, M.: Domain Adaptation of Recurrent Neural Networks for Natural Language Understanding, in *Proceedings of INTERSPEECH*, pp. 690–694 (2016)

[Jiang 21] Jiang, T., Wang, D., Sun, L., Yang, H., Zhao, Z., and Zhuang, F.: LightXML: Transformer with Dynamic Negative Sampling for High-Performance Extreme Multi-label Text Classification, *arXiv preprint arXiv:2101.03305* (2021)

[Kayama 10] Kayama, K., Kobayashi, A., Mizukami, E., Misu, T., Kashioka, H., Kawai, H., and Nakamura, S.: Spoken Dialog System on Plasma Display Panel Estimating Users' Interest by Image Processing., in *Proceedings of the 6th International Conference on Intelligent Environments*, pp. 4–13 (2010)

[Kobayashi 18] Kobayashi, Y., Yoshida, T., Iwata, K., Fujimura, H., and Akamine, M.: Out-of-Domain Slot Value Detection for Spoken Dialogue Systems with Context Information, in *Proceedings of 2018 IEEE Spoken Language Technology Workshop (SLT)*, pp. 854–861 (2018)

[Kobayashi 19] Kobayashi, Y., Yoshida, T., Iwata, K., and Fujimura, H.: Slot Filling with Weighted Multi-Encoders for Out-of-Domain Values, in *Proceedings of INTERSPEECH*, pp. 854–858 (2019)

[Kolachina 17] Kolachina, P., Riedl, M., and Biemann, C.: Replacing OOV Words For Dependency Parsing With Distributional Semantics, in *Proceedings of the 21st Nordic Conference on Computational Linguistics*, pp. 11–19 (2017)

[Lafferty 01] Lafferty, J., Andrew, M., and Pereira, F. C. N.: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, in *Proceedings of the 18th International Conference on Machine Learning*, pp. 282–289 (2001)

[Lee 21] Lee, J., Sung, M., Kang, J., and Chen, D.: Learning Dense Representations of Phrases at Scale, in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 6634–6647, Online (2021), Association for Computational Linguistics

[Liu 16] Liu, B. and Lane, I.: Attention-Based Recurrent Neural Network Models for Joint Intent Detection and Slot Filling, in *Proceedings of INTERSPEECH*, pp. 685–689 (2016)

[Liu 20] Liu, Z., Winata, G. I., Xu, P., and Fung, P.: Coach: A Coarse-to-Fine Approach for Cross-domain Slot Filling, in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 19–25 (2020)

[Liu 21] Liu, L., Lin, X., Zhang, P., and Wang, B.: Improving Cross-Domain Slot Filling with Common Syntactic Structure, in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7638–7642 (2021)

[Louvan 20] Louvan, S. and Magnini, B.: Simple is Better! Lightweight Data Augmentation for Low Resource Slot Filling and Intent Classification, in *Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation*, pp. 167–177, Hanoi, Vietnam (2020), Association for Computational Linguistics

[Luong 15] Luong, T., Sutskever, I., Le, Q., Vinyals, O., and Zaremba, W.: Addressing the Rare Word Problem in Neural Machine Translation, in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pp. 11–19 (2015)

[Makazhanov 16] Makazhanov, A. and Yessenbayev, Z.: Character-based feature extraction with LSTM networks for POS-tagging task, in *Proceedings of 2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT)*, pp. 1–5 (2016)

[McCallum 00] McCallum, A., Freitag, D., and Pereira, F. C. N.: Maximum Entropy Markov Models for Information Extraction and Segmentation, in *Proceedings of the 17th International Conference*

on Machine Learning, pp. 591–598, San Francisco, CA, USA (2000)

[Mikolov 13] Mikolov, T., Chen, K., Corrado, G., and Dean, J.: Efficient Estimation of Word Representations in Vector Space, in *Proceedings of the International Conference on Learning Representations (ICLR)*, pp. 1310–1318 (2013)

[Nakayama 18] Nakayama, H.: seqeval: A Python framework for sequence labeling evaluation (2018), Software available from https://github.com/chakki-works/seqeval

[Nisimura 05] Nisimura, R., Lee, A., Yamada, M., and Shikano, K.: Operating a Public Spoken Guidance System in Real Environment, in *Proceedings of INTERSPEECH*, pp. 845–848 (2005)

[Ouchi 18] Ouchi, H., Shindo, H., and Matsumoto, Y.: A Span Selection Model for Semantic Role Labeling, in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 1630–1642, Brussels, Belgium (2018), Association for Computational Linguistics

[Pennington 14] Pennington, J., Socher, R., and Manning, C.: Glove: Global Vectors for Word Representation, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543 (2014)

[Prange 15] Prange, J., Thater, S., and Horbach, A.: Unsupervised Induction of Part-of-Speech Information for OOV Words in German Internet Forum Posts, in *Proceedings of the 2nd Workshop on Natural Language Processing for Computer-Mediated Communication / Social Media (NLP4CMC)* (2015)

[Rajendran 18] Rajendran, J., Ganhotra, J., Guo, X., Yu, M., and Singh, S.: Named Entities troubling your Neural Methods? Build NE-Table: A neural approach for handling Named Entities, *arXiv preprint arXiv:1804.09540v1* (2018)

[Raux 05] Raux, A., Langner, B., Bohus, D., Black, A., and Eskenazi, M.: Let's Go Public! Taking a Spoken Dialog System to the Real World, in *Proceedings of INTERSPEECH*, pp. 885–888 (2005)

[Raymond 07] Raymond, C. and Riccardi, G.: Generative and Discriminative Algorithms for Spoken Language Understanding, in *Proceedings of INTERSPEECH*, pp. 1605–1608 (2007)

[Ribeiro 18] Ribeiro, E., Ribeiro, R., and Matos, de D. M.: A Study on Dialog Act Recognition Using Character-Level Tokenization, in Agre, G., Genabith, van J., and Declerck, T. eds., *Artificial Intelligence: Methodology, Systems, and Applications*, pp. 93–103, Springer International Publishing (2018)

[Schuster 97] Schuster, M. and Paliwal, K. K.: Bidirectional Recurrent Neural Networks, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 45, No. 11, pp. 2673–2681 (1997)

[Seneff 92] Seneff, S.: Robust Parsing for Spoken Language Systems, in *Proceedings of 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. 1, pp. 189–192 (1992)

[Shah 18] Shah, P., Hakkani-Tür, D., Tür, G., Rastogi, A., Bapna, A., Nayak, N., and Heck, L.: Building a Conversational Agent Overnight with Dialogue Self-Play, *arXiv preprint arXiv:1801.04871* (2018)

[Shin 18] Shin, Y., Yoo, K. M., and Lee, goo S.: Slot Filling with Delexicalized Sentence Generation, in *Proceedings of INTERSPEECH*, pp. 2082–2086 (2018)

[Siddique 21] Siddique, A., Jamour, F., and Hristidis, V.: Linguistically-Enriched and Context-Aware Zero-Shot Slot Filling, in *Proceedings of the Web Conference 2021*, *WWW '21*, pp. 3279–3290, New York, NY, USA (2021), Association for Computing Machinery

[Simonnet 15] Simonnet, E., Deleglise, P., Camelin, N., and Esteve, Y.: Exploring the Use of Attention-based Recurrent Neural Networks for Spoken Language Understanding, in *Proceedings of Machine Learning for SLU and Interaction NIPS 2015 Workshop (SLU-NIPS 2015)* (2015)

[Sutskever 14] Sutskever, I., Vinyals, O., and Le, Q. V.: Sequence to Sequence Learning with Neural Networks, in *Proceedings of Advances in Neural Information Processing Systems 27*, pp. 3103–3112 (2014)

[Wang 18] Wang, G., Li, C., Wang, W., Zhang, Y., Shen, D., Zhang, X., Henao, R., and Carin, L.: Joint Embedding of Words and Labels for Text Classification, in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pp. 2321–2331 (2018)

[Wang 20] Wang, Y., Shen, Y., and Jin, H.: A BI-Model Approach for Handling Unknown Slot Values in Dialogue State Tracking, in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8019–8023 (2020)

[Wolf 20] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., Platen, von P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., Drame, M., Lhoest, Q., and Rush, A. M.: Transformers: State-of-the-Art Natural Language Processing, in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 38–45, Online (2020), Association for Computational Linguistics

[Wu 19] Wu, C.-S., Madotto, A., Hosseini-Asl, E., Xiong, C., Socher, R., and Fung, P.: Transferable Multi-Domain State Generator for Task-Oriented Dialogue Systems, in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 808–819, Florence, Italy (2019), Association for Computational Linguistics

[Xu 18] Xu, P. and Hu, Q.: An End-to-end Approach for Handling Unknown Slot Values in Dialogue State Tracking, in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 1448–1457 (2018)

[Young 13] Young, S., Gasic, M., Thomson, B., and Williams, J. D.: POMDP-Based Statistical Spoken Dialog Systems: A Review, *Proceedings of the IEEE*, Vol. 101, No. 5, pp. 1160–1179 (2013)

[Zhang 17] Zhang, Y., Zhong, V., Chen, D., Angeli, G., and Manning, C. D.: Position-aware Attention and Supervised Data Improve Slot Filling, in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 35–45 (2017)

[Zhao 18] Zhao, L. and Feng, Z.: Improving Slot Filling in Spoken Language Understanding with Joint Pointer and Attention, in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pp. 426–431 (2018)

[Zhong 18] Zhong, V., Xiong, C., and Socher, R.: Global-Locally Self-Attentive Encoder for Dialogue State Tracking, in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pp. 1458–1467 (2018)

〔担当委員：北岡 教英〕

2021 年 9 月 7 日　受理

◇ 付　　録 ◇

A. 実験結果の条件

表 A.1　実験結果の条件

データセット	モデル	倍率	WD	TFD
Sim-M	VD+PS+ML	×1	0.0	0.3
Sim-M	SE+CWE+PNS	×1	0.5	0.0
Sim-R	VD+PS	×1	0.0	0.0
Sim-R	SE+PNS	×5	0.3	0.0
Shop	VD+PS+ML	×3	0.3	0.0
Shop	SE+CWE+PS	×5	0.5	0.0

(倍率:PS/NS/PNS の倍率, WD:Word dropout ,
TFD:Target feature dropout)

表 A.2　ポジティブ・ネガティブサンプルの倍率による性能比較

(a) バリュー検出器　再現率

データセット	モデル	PS/NS/PNS 倍率		
		×1	×3	×5
Sim-M	VD+PS+ML	**96.50**	96.00	96.27
Sim-R	VD+PS	**96.18**	95.84	95.63
Shop	VD+PS+ML	89.65	**90.08**	89.38

(b) スロット推定器　F 値

データセット	モデル	PS/NS/PNS 倍率		
		×1	×3	×5
Sim-M	SE+CWE+PNS	**96.66**	96.59	96.63
Sim-R	SE+PNS	96.28	96.29	**96.32**
Shop	SE+CWE+PS	90.30	90.28	**90.67**

――――著　者　紹　介――――

小林　優佳（正会員）

2004 年東京工業大学大学院理工学研究科機械制御システム専攻修士課程修了．同年，東芝家電製造株式会社 (現東芝ホームアプライアンス株式会社) 入社．2008 年株式会社東芝入社．音声対話の研究に従事．電子情報通信学会，情報処理学会，各会員．

久島　務嗣

2020 年電気通信大学大学院情報理工学研究科情報学専攻修士課程修了．同年，株式会社東芝入社．音声対話の研究に従事．電子情報通信学会，会員．

吉田　尚水

2013 年東京工業大学大学院情報理工学研究科情報環境学専攻修了．博士（工学）．同年，株式会社東芝入社．音声信号処理，生体信号処理，音声対話システムの研究開発に従事．

藤村　浩司

2005 年名古屋大学大学院情報科学研究科メディア科学専攻修士課程修了．同年，株式会社東芝入社．音声認識・対話の研究に従事．電子情報通信学会，音響学会，各会員．

岩田　憲治

2008 年東京工業大学大学院情報理工学研究科情報工学専攻修士課程修了．同年，株式会社東芝入社．音声対話の研究に従事．