

# [ポスター講演] Attention-based LSTM を用いた意図理解とキーワード抽出の統合による質問応答システム

松好 祐紀<sup>†</sup> 滝口 哲也<sup>†</sup> 有木 康雄<sup>†</sup>

<sup>†</sup> 神戸大学大学院システム情報学研究科

〒657-8501 兵庫県神戸市灘区六甲台町 1-1

E-mail: †yuki.matsuyoshi@stu.kobe-u.ac.jp, ††{takigu,ariki}@kobe-u.ac.jp

あらまし 情報社会では、あらゆる場面で機械やソフトウェアを使う機会が多い。そういったシステムを使う前に、最初にマニュアルを読むが、それだけで使い慣れるのは困難である。本研究では、ユーザーがデバイスやソフトウェアなどのシステムを使う上で生じる質問に、システム自体が対話的に答えていくことで、人間の理解を助けて習熟を可能にするシステムの構築を目的としており、その前段階として、オセロゲームを対象とした対話的な質問応答形式でのサポートを考える。システムがユーザーを満足させる回答をするには、ユーザーの質問意図を正確に理解する必要がある。そこで本稿では、ユーザーの質問の意図として、ユーザーの質問文から質問タイプ、質問キーワードクラスを、*attention* 機構を導入した LSTM モデルを使用して推定した。また、質問タイプ、質問キーワードクラスの推定モデルを結合して、一つのモデルで推定を行うことが出来る *joint* 学習モデルを構築した。

キーワード 質問応答システム、LSTM、Attention 機構、Joint 学習

## Question answering system by integration of intention understanding and keyword extraction using attention-based LSTM

Yuki MATSUYOSHI<sup>†</sup>, Tetsuya TAKIGUCHI<sup>†</sup>, and Yasuo ARIKI<sup>†</sup>

<sup>†</sup> Graduate School of System Informatics, Kobe University

Rokkodai 1-1, Nada-ku, Kobe-shi, Hyogo, 657-8501 Japan

E-mail: †yuki.matsuyoshi@stu.kobe-u.ac.jp, ††{takigu,ariki}@kobe-u.ac.jp

**Abstract** In an information society, we have many opportunities to use machines and computer software in various fields. Before using these technology, we read the operation manuals at the beginning, but it is difficult for us to use them as we want. From this view point, in this study, we aim to construct a system that can help users use them by answering their questions interactively during the operations. As a preliminary step, an interactive support system is constructed for Othello games. In order for the system to answer user's question satisfactory, it is necessary to accurately understand the user's intention from the question. In this paper, for that purpose, we propose to estimate *question type* and *question keyword class* from user's question sentences using the LSTM model with *attention* mechanism. We also propose *joint* models for simultaneous estimation of *question type* and *question keyword class*.

**Key words** question answeing system, LSTM, attention mechanism, joint learning

### 1. はじめに

情報社会である現代では、職種を問わず機械やプログラムなどのシステムを使う機会が多い。最初はマニュアルを読むことになるが、それだけで使い慣れるのは困難であり、実際に使いながら慣れていくことが多い。この際に熟練者のサポートを得る

ことが出来れば、習熟が早い。そこで、本研究では、ユーザーがシステムを使う上で生じる質問に、システム自体が対話的に答えていくことで、人間の理解を助けて習熟を可能にするシステムの構築を目的としており、その前段階として、オセロゲームを対象とした対話的なサポートを考えている。これは、オセロゲームを始めとしたボードゲームもルールブックを読むだけ

ではなく、実際にプレイしながら上手くなっていくものだと考えたからである。

一般的に、自然言語理解システムは、ユーザーの発話からユーザーの意図と意味構造を抽出するために、*intent detection* (意図検出) と *slot filling* を行う。*Intent detection* は発話の分類問題、*slot filling* はシーケンスラベリング問題として考えることが出来る。*Intent detection* では、SVM [1], ディープニューラルネットワーク (DNN) [2] が、*slot filling* では、リカレントニューラルネットワーク (RNN) [3], 畳み込みニューラルネットワーク (CNN) [4] がよく利用されている。最近では、*encoder-decoder* ニューラルネットワークが *slot filling* でよく利用されるようになってきている [5]。さらに、*attention* 構造 [6] を利用することで、*encoder-decoder* モデルが、デコードと入力データのアライメントを同時に学習出来るようになる。

本稿では、質問応答の形式での対話的なサポートを考えており、ユーザーの質問文から質問タイプと質問キーワードクラスを推定することで、ユーザーの質問意図を特定する。質問タイプ推定は *intent detection*, 質問キーワードクラス推定は *slot filling* であると見なすことが出来る。我々は、質問タイプの推定に用いる *attention-based Long Short Term Memory (LSTM)* モデル、質問キーワードクラスの推定に用いる *attention-based LSTM encoder-decoder* モデルを提案する。また、質問タイプの推定モデルと質問キーワードクラスの推定モデルを結合し、両方を同時に推定出来る、*joint* モデルを提案する。

## 2. これまでの研究の問題点

我々は以前、質問応答システムをルールベースで構築した。このシステムは、質問解析部と回答生成部で構成されている。質問解析部では、ユーザーの質問意図として、ユーザーの質問文から質問タイプと質問キーワードクラスが、キーワードスポッティングで抽出される。質問タイプの数は5つ、質問キーワードクラスは21個とした。質問タイプと質問キーワードクラスは、オセロプログラムからのパラメータと合わせて回答生成部に渡される。

回答生成部では、質問解析部から渡されるパラメータを条件とそれらに対応する回答テンプレートから構成される。条件によって選択された回答テンプレートに質問キーワードクラスの値 (質問キーワード) と、オセロプログラムからのパラメータを埋め込むことで、システムの回答を生成する。ルールは202種類作成した。

実際に構築したシステムの評価を行ったところ、様々な問題が生じた。実際のプレイ中になされた質問を分析した結果、想定以上に多様であり、if-then ルールで質問を解析するには限界があることが分かった。また、将来的に他のドメインも扱えるようにする事を考えており、ルールベースの場合、ドメイン毎にルールを作らなければならなくなる。以上の問題を解決するには、ドメイン知識を与えるだけで、システムが自動で質問を解析できる必要があると考えた。

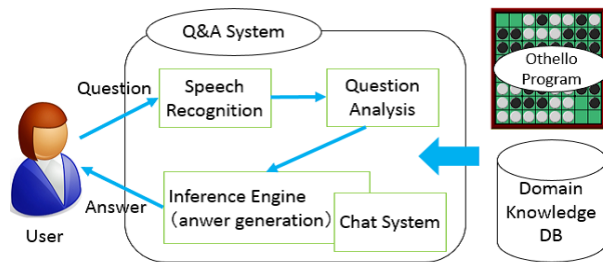


図1 システムの概要.

Fig. 1 Outline of our system.

## 3. 提案システム

### 3.1 システムの概要

我々のシステムの概要を図1に示す。まず、ユーザーの質問文が質問解析部に渡される。そして、質問解析の結果とオセロプログラムからのパラメータ、オセロの知識データベースの情報を用いて、推論エンジンにより最適な回答を生成する。また、質問解析部において質問文が雑談である、と判断されたら推論エンジンの代わりに雑談対話システムが回答を生成する。本稿では、質問解析部の構築に焦点を当てる。

### 3.2 質問解析部

本研究では日本語を取り扱うため、日本語形態素解析器 MeCab を利用して、ユーザーの質問文を形態素列に分解している。質問タイプと質問キーワードクラスは、ユーザーの質問文の形態素列から推定される。質問タイプは質問の概要であり、ユーザーの大まかな質問意図を表す。質問タイプとして、これまでの研究のものを見直し、合計15種類を定義した。以下に例を示す。

- 理由: システムの回答などに関して生じる疑問に関する質問
  - 場所: オセロ盤面上の位置に関する質問
  - 雑談: オセロに関する質問以外の一般的な会話
- 質問キーワードクラスは、質問文中に現れるキーワードの上位語であり、合計14種類定義している。以下に例を示す。
- 用語: X 打ち, 開放度など, オセロの専門用語
  - 座標: b7,a8 など, オセロ盤面上の位置の呼び方

### 3.3 質問タイプ推定

質問タイプは、図2に示すような *attention-based bidirectional LSTM* [7] で推定される。このモデルは、LSTM モデルに *attention* 機構を導入したものである。まず、質問文中の各形態素  $x_i (i=1,2,\dots,m)$  が one-hot 単語ベクトルに変換される。単語ベクトルは、word embedding によって分散表現に変換され、時系列的にモデルに入力される。bidirectional LSTM の順方向の隠れ層  $h_{fi}$  と逆方向の隠れ層  $h_{bi}$  は、学習パラメータ  $W_1, W_2$  を用いて連結し、 $h_i$  にする。

$$h_i = W_1 h_{fi} \oplus W_2 h_{bi} (i = 1, 2, \dots, m) \quad (1)$$

文の終わりを表す  $\langle \text{eos} \rangle$  が入力された時、モデルはその時点での LSTM の隠れ層  $h_o$  を出力する。*attention* 重み  $\alpha_i$  は、 $\langle \text{eos} \rangle$  以前の入力  $h_i (i=1,2,\dots,m)$  と  $h_o$  から、以下

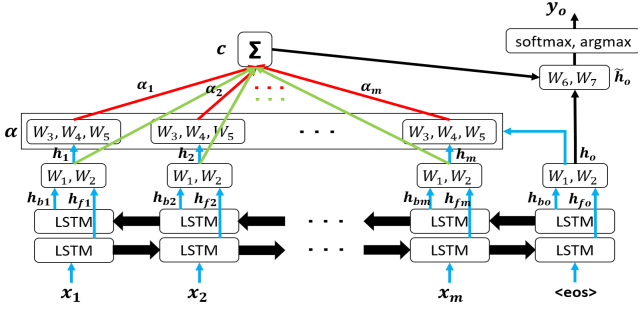


図2 質問タイプ推定を行う *attention-based bidirectional LSTM* モデル。

Fig.2 *Attention-based LSTM* model for *question type* estimation.

のように計算される。

$$\alpha_i = \frac{\exp(W_3^T \tanh(W_4 h_i + W_5 h_o))}{\sum_{j=1}^m \exp(W_3^T \tanh(W_4 h_j + W_5 h_o))} \quad (2)$$

$W_3, W_4, W_5$  は学習パラメータである。  $\alpha_i$  を  $h_i$  の重みとし、コンテキストベクトル  $c$  を計算する。

$$c = \sum_{i=1}^m \alpha_i h_i \quad (3)$$

$c$  と  $h_o$  から、以下の式により  $\tilde{h}_o$  を計算する。

$$\tilde{h}_o = \tanh(W_6 c + W_7 h_o) \quad (4)$$

$W_6, W_7$  は学習パラメータである。  $\tilde{h}_o$  のサイズは、出力語彙(質問タイプの数で、15次元)になる。  $\tilde{h}_o$  に softmax 関数を用いさせ、その中の最大の値を質問タイプ  $y_o$  の推定値として選択する。

$$y_o = \operatorname{argmax}(\operatorname{softmax}(\tilde{h}_o)) \quad (5)$$

### 3.4 質問キーワードクラス推定

質問キーワードクラス推定は、*slot filling* として扱う事ができる。 *slot filling* では、入力のシーケンスと、ラベル(出力)のシーケンスが一致するようにモデルを学習する。質問キーワードクラス推定では、ユーザーの質問文の形態素列である入力シーケンスと、それに対応する質問キーワードクラスのラベルシーケンスの長さは同じになる。この際、質問キーワードではない形態素(例えば、「は」、「が」などの助詞や句読点など)はアルファベットの”O”で置き換える。例を図3に示す。

質問キーワードクラスは、図4に示すような *attention-based LSTM encoder-decoder* モデル [7,8] で推定される。このモデルは、従来の LSTM encoder-decoder [4] に *attention* 機構を導入したものである。エンコーダは 3.3 で述べた *attention-based bidirectional LSTM* と同様である。デコーダは、モデルに  $\langle \text{eos} \rangle$  が入力された後、質問キーワードクラス ( $y_1, y_2, \dots$ ) が順番に生成されるように学習される。

従来の LSTM encoder-decoder モデル [4,8] では、エンコーダに質問文の形態素列  $x_i (i=1,2,\dots,m)$  が入力され、それらの情報が隠れ層  $h_o$  に集約される。よって、この  $h_o$  は、入力質問文全体の情報を持っており、デコーダにおいて出力シーケンスである質問キーワードクラス列  $y$  の生成に利用される。 *attention*

形態素	何で	“h8”	が	良い	手	?
質問キーワードクラス	0	座標	0	形容詞良	手	0
質問タイプ	理由					

図3 本研究でのコーパス(質問文、質問タイプ、質問キーワードクラス)の例。

Fig.3 Sample of corpus in this study with *question type* and *question keyword class* annotation.

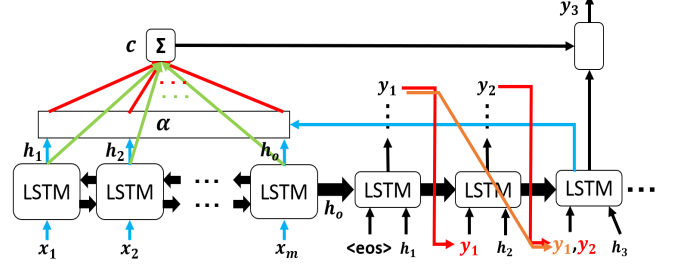


図4 質問キーワードクラス推定を行う *attention-based LSTM encoder-decoder* モデル。

Fig.4 *Attention-based LSTM encoder-decoder* model for *question keyword class* estimation.

機構のコンテキストベクトル  $c$ 、及び *attention* 重み  $\alpha$  は、デコーダにおいて出力の各時間ステップごとに計算される。デコーダは、以下のような、出力シーケンス  $y$  の確率を計算する。

$$P(y | h_o) = \prod_{t=1}^T P(y_t | y_1^{t-1}, h_o) \\ \equiv \prod_{t=1}^T P(y_t | y_{t-1}, h_o) \quad (6)$$

ここで  $y_{t-1}$  は、現在の時間ステップ  $t$  から一つ前の時間ステップ  $t-1$  の隠れ層の出力であり、  $y_1^{t-1}$  は  $y_1, y_2, \dots, y_{t-1}$  を表している。

提案モデルでは、デコーダは現在の時間ステップ  $t$  の出力  $y_t$  の確率の計算に、2つ前の時間ステップ  $t-2$  の隠れ層の出力  $y_{t-2}$  も利用する。シーケンスラベリングでは、より長い文章を扱うことで結果が向上するので、以下の式 (7) は式 (6) と比べて条件付き確率の向上に寄与する。

$$P(y | h_o) = \prod_{t=1}^T P(y_t | y_1^{t-1}, h_o) \\ \equiv \prod_{t=1}^T P(y_t | y_{t-1}, y_{t-2}, h_o) \quad (7)$$

### 3.5 Joint モデル

*Intent detection* と *slot filling* の *joint* モデルは [8] や [9] で提案されている。このようなモデルでは、 *intent detection* と *slot filling* を一つのモデルで学習することが出来る。本研究では、 3.3 で述べた質問タイプの推定モデルと、 3.4 で述べた質問キーワードクラスの推定モデルを結合した *joint* モデルを6種類 (*model1, model2, model3, model4, model5, model6*) 構築した。

*Model1* の概略図を図5に示す。エンコード部は共通であり、デコード部が質問タイプ推定のデコーダと質問キーワードクラス推定のデコーダに分かれている。モデルを学習する際、両方

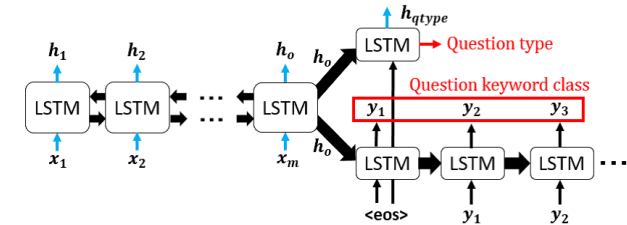


図5 model1 の概略図.

Fig. 5 Schematic of model1

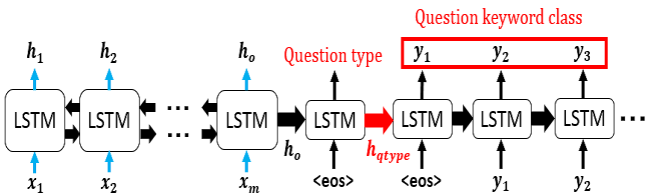


図6 model2 の概略図.

Fig. 6 Schematic of model1

のデコーダの誤差値がエンコーダに逆伝播される。質問タイプ推定のデコーダと質問キーワードクラス推定のデコーダはエンコーダの最終の隠れ層の出力  $h_o$  を共有しており、これは 3.4 で述べたとおり、質問文全体の情報をエンコードしている。また、図5では省略したが、*attention* 機構を導入しており、エンコード部で計算したアテンションベクトル  $\alpha$  は両方のデコーダで共有している。

Model2 の概略図を図6に示す。エンコード部はモデル1と同様に共通であるが、デコード部において、質問タイプ推定のデコードを先に行い、質問キーワードクラス推定のデコーダには、エンコーダの最終の隠れ層の出力  $h_o$  ではなく、質問タイプ推定のデコーダの隠れ層の出力  $h_{qtype}$  が渡される。このモデルでも、*attention* 機構を導入している。

Model3,4 は、それぞれ model1,2 に関して、デコード部の各時間ステップ  $t$  で、モデルへの入力に質問タイプ推定のデコーダの隠れ層の出力  $h_{qtype}$  も加えたモデルである。

$$h_t = LSTM(Ey_{t-1}, [h_{t-1}, h_{t-2}, h_{qtype}]) \quad (8)$$

$h_t$  はデコーダの時間ステップ  $t$  での隠れ層の出力、 $y_{t-1}$  は一つ前の時間ステップの質問キーワードクラス、 $E$  は埋め込み行列である。

Model5,6 は、それぞれ model1,2 に関して、デコード部の各時間ステップ  $t$  で、モデルへの入力に、 $h_{qtype}$  に *attention* 機構のコンテキストベクトル  $c_t$  も加味した出力  $h_{ct}$  (式(4)参照)を加えたモデルである。

$$h_t = LSTM(Ey_{t-1}, [h_{t-1}, h_{t-2}, h_{ct}]) \quad (9)$$

$$(h_{ct} = \tanh([c_t, h_{qtype}]))$$

## 4. 実験

### 4.1 実験の詳細

データセットとしては、これまでの研究で構築したルールベースの質問応答システムを試用してもらった際に収集した、

表1 データセット中の質問タイプの分布.

Table 1 The distribution of *question type* in the training data.

理由1	理由2	場所1	場所2	場所3	定義	結果	確認1
170	104	93	150	104	152	97	139
確認2	確認3	提案	選択	常識	戦略	雑談	
86	110	153	114	99	90	127	

表2 データセット中の質問キーワードクラスの分布.

Table 2 The distribution of *question keyword class* in the training data.

座標1	座標2	形容詞良	形容詞悪	手番	否定	(石の)増加
396	415	422	205	176	122	134
用語	手	勝敗	評価	石数	比較	指示語
438	423	116	173	115	124	144

表3 質問タイプ推定の実験結果.

Table 3 Experimental results of *question type* estimation.

モデル	推定率 (%)	再現率	適合率
Rule-based	75.8	-	-
SVM	82.6	0.80	0.82
Random Forest	77.5	0.75	0.79
Non-attention			
bidirectional LSTM	84.2	0.82	0.78
Attention-based			
bidirectional LSTM	85.2	0.81	0.83

オセロゲームをプレイする際に生じるユーザー質問文データのコーパスを用いる。このコーパスでは、1つの質問文に対し、質問タイプ、質問キーワードクラスがアノテーションされている。このアノテーションは人手で行っている。学習データの総数は1788文であり、そのうちの1605文を学習データ、183文をテストデータとして用いた。

データセット中の質問タイプと質問キーワードクラスの分布を表1、表2に示す。表1の「理由1」、「理由2」などは、質問タイプを更に分割したものである。今回のデータセットでは、1質問文中に平均して1.9個の質問キーワードが含まれている。

LSTMでは、word embeddingsはランダムに初期化され、サイズは250次元である。ドロップアウトは使用しなかった。また、モデルの最適化にはAdamを使用した。Adamのパラメータは ( $\alpha = 0.0001$ ,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ,  $\epsilon = 10^{-8}$ ) となっている。これらのハイパーパラメータは、グリッドサーチにより決定した。

### 4.2 質問タイプ推定の結果

質問タイプ推定では、テストデータが入力されたモデルの出力から正しく推定できている数を調べ、推定率を計算することでモデルを評価する。3.3で述べた *attention-based bidirectional LSTM* モデルに加え、比較のために、*attention* 機構を導入しない *non-attention bidirectional LSTM* モデル、以前の研究で構築したルールベースモデル、SVM、Random Forestでも同様の実験を行った。

表3に実験の結果を示す。*Bidirectional LSTM* で質問タイプの推定を行った場合、*attention* 機構を導入した場合は、推定率が84.7%、導入しない場合は83.7%となり、ルールベー

表 4 質問キーワードクラス推定の実験結果.

Table 4 Experimental results of *question keyword class estimation*.

モデル	推定率 (%)
Non-attention LSTM Encoder-Decoder[4]	78.7
Attention-based LSTM Encoder-Decoder[8]	79.2
Our proposed model	81.4

システム, SVM, Random Forest での推定結果を上回った. *bidirectional LSTM* モデルの結果としては, *attention* 機構を導入することによる推定率の向上が, 期待していたようにはならなかった. 原因としては, 質問タイプ推定は単純な推定問題になるので, *bidirectional LSTM* だけで十分高い推定率を出す事ができ, *attention* 機構は不要であったということが分かった.

### 4.3 質問キーワードクラス推定の結果

質問キーワードクラス推定でも, モデルの推定率を計算して評価する. 推定率は, テストデータを入力したモデルの出力シーケンスの中に, 正しい質問キーワードクラスが全て含まれている場合を数えて計算する. 3.4 で述べた, デコーダ部において現在の時間ステップのモデルへの入力に, 2つ前の時間ステップの出力も加える, 提案モデルである *attention-based bidirectional LSTM encoder-decoder* モデルに加え, 比較のために, *attention* 機構を導入しない *non-attention LSTM encoder-decoder* モデル [4], 従来の *attention-based LSTM encoder-decoder* モデル [8](デコーダ部において現在の時間ステップのモデルへの入力に, 1つ前の時間ステップの出力のみ利用) でも同様の実験を行った.

表 4 に実験の結果を示す. 提案モデルは, 他のモデルと比べて高い推定率を出した. これにより, *attention* 構造と提案手法が有効であることが示された. 実際にモデルからの出力を調べたところ, 入力の質問文中に質問キーワードが3個以上含まれる場合は, 3個目以降が上手く推定できていないことが多かった. 原因の一つとして, 4.1 でも述べたが, データセット全体で, 1質問文中に平均して 1.9 個の質問キーワードしか含まれていないので, 3個以上含まれる場合を上手く学習できていないことが考えられる. また, 推定が上手く行かなかった場合「座標 1」と「座標 2」, 「形容詞量」と「形容詞良」など, 似たような意味の質問キーワードクラスが間違っ推定されることが多かった.

### 4.4 Joint モデル推定の結果

3.5 でも述べた通り, 6つの *joint* モデルで実験を行い, 結果を比較した. 表 5 に実験の結果を示す. 結果を見ると, 質問タイプ, 質問キーワードクラス両方で最も高い推定率を出したモデルはなかった. *Model2* と *model6* が全体としては良い結果を出していることが分かる. しかし, 質問タイプ推定に関しては, *model2*, *model3* の推定率が高かったが, 4.2 の一番良いモデルの結果より 0.5% 下がった. 質問キーワードクラス推定に関しては, *model2,4,6* の推定率が高かったので, *joint* モデルとしては, 図 6 の形の方が良いという結果になった. ま

表 5 *joint* モデル推定の質問タイプ推定に関する結果.Table 5 Experimental results of *joint* model estimation.

モデル	質問タイプ			質問キーワード クラスの推定率 (%)
	推定率 (%)	再現率	適合率	
<i>model1</i>	83.1	0.82	0.73	81.4
<i>model2</i>	84.7	0.84	0.75	83.6
<i>model3</i>	84.7	0.8	0.81	81.4
<i>model4</i>	83.6	0.82	0.76	83.1
<i>model5</i>	84.2	0.83	0.74	82.5
<i>model6</i>	84.2	0.80	0.79	84.2

表 6 質問タイプ推定モデルと質問キーワードクラス推定モデルの推定結果のエントロピー. (エントロピーが小さいほど, モデルの信頼度が高い.)

Table 6 Entropy of *question type* estimation model and *question keyword class* estimation model. (The smaller the entropy, the higher the reliability of the model.)

	質問タイプ推定モデル	質問キーワードクラス推定モデル
信頼度	0.00026	0.021

た, デコード部の各時間ステップ  $t$  で, モデルへの入力に質問タイプ推定のデコーダの隠れ層の出力  $h_{qtype}$  を加える試み (式 (8)) に関しては, *model4* が *model2* より結果が悪いことを考えると, 良い手法ではなかったと思われるが,  $h_{qtype}$  の代わりに  $h_{ct}$  を加えた (式 (9)) *model6* に関しては, 質問キーワードクラスの推定率が一番高くなったので, 一定の効果が見られた.

*Joint* モデルでは, 質問タイプ推定の推定率が, 単独で推定した場合と比べて低下した. この原因を調べるために, 質問タイプ推定モデル (4.2 の *attention-based bidirectional LSTM* モデル) と質問キーワードクラス推定モデル (??の Our proposed model) による推定結果の信頼度を調べた. 各モデルの推定結果  $y$  は, 3.3 で述べたように, モデルの出力  $h$  から式 (5) のように計算される.

$$y = \operatorname{argmax}(\operatorname{softmax}(h)) \quad (5)$$

$h$  に関して  $\operatorname{softmax}$  を適用した  $\operatorname{softmax}(h)$  はベクトルであり, 質問タイプ推定モデルの場合は, 質問タイプの種類の数である 15 次元, 質問キーワードクラス推定モデルの場合は, 質問キーワードクラスの種類の数である 14 次元に, 質問キーワードに値しない形態素に付けられるラベル "O" を加えた 15 次元となっている. これらのベクトルのエントロピーを算出した.  $s = \operatorname{softmax}(h)$  とすると, エントロピーの式は以下ようになる.

$$\operatorname{entropy} = - \sum_{i=1}^n s_i \log_2 s_i \quad (10)$$

この値が小さいほど, 推定結果の信頼度が高いということになる. 実際に, 各モデルの信頼度を計算した結果, 表 6 の様になった. 質問キーワードクラス推定に関しては, テストデータ 1 文に対して複数の出力があるため, それらのエントロピーを平均している.

表 6 より, 質問タイプ推定モデルの方が, 質問キーワードクラス推定モデルよりも信頼度が高いことが分かる. この結果か

ら, *joint* 学習を行うことで, 質問タイプ推定の信頼性により質問キーワードクラス推定の推定精度が向上し, 逆に, 質問キーワードクラス推定の信頼性の低さにより, 質問タイプの推定精度が低下したのではないかと考えられる. よって, 質問キーワードクラス推定を行う場合は *joint* モデルを, 質問タイプ推定を行う場合は単独のモデルを用いた方がよい, ということになる.

## 5. おわりに

本稿では, ユーザーの質問意図を把握するためにユーザーの質問の質問タイプと質問キーワードクラスを推定するモデルを提案した. 質問タイプ推定では, *attention-based bidirectional LSTM* モデルが最も高い推定率を出したが, *attention* 機構を導入しない *bidirectional LSTM* モデルの推定率とほぼ変わらなかったため, *attention* 機構を導入することによる顕著な効果を確認出来なかった. 質問キーワードクラス推定では, *attention* 機構と, 提案手法の導入によって, 従来の *bidirectional LSTM encoder-decoder* モデルよりも結果が向上した. *Joint* モデル推定では, 質問タイプと質問キーワードクラスの推定モデルの結合の仕方を変えた, 6つのモデルを構築し, 比較を行った. 結果としては, *model2* による質問タイプの推定率は, 結合前と比べて0.5%低下したが, 質問キーワードクラスの推定率は結合前と比べて2.2%向上した. また, *model6* は, 質問タイプの推定率が, 結合前と比べて1%低下したが, 質問キーワードクラスの推定率が2.8%向上した. この結果から, 結合の仕方としては, 図6が良いという結果になった. また, デコード部においてモデルへの入力に  $h_{ct}$  を加える (3.5) ことによって, 質問キーワードクラスの推定率が向上した. 結論として, 質問タイプ推定を行う場合は単独のモデルを, 質問キーワードクラス推定を行う場合は *joint* 学習モデルを用いた方がよい, ということが分かった. これは, *joint* 学習をする時の各モデルの信頼度に依存すると考えられる.

今後は, 今回の *joint* 推定モデルに, ドメイン特有の辞書情報を加えることによってさらに推定率を向上させることを考えている. また, *attention* 機構がどれだけ有効であるかを, *attention* 重みベクトルを調べることで解析することを試みる.

謝辞 本研究の一部は, JSPS 科研費 JP17K00236 の助成を受けたものである.

## 文 献

- [1] P. Haffner, G. Tur, and J. H. Wright, "Optimizing svms for complex call classification," in *Acoustics, Speech, and Signal Processing, 2003, Proceedings. (ICASSP'03). 2003 IEEE International Conference on*, vol. 1, IEEE, 2003, pp. 529-551.
- [2] R. Sarikaya, G. E. Hinton, and B. Ramadhadran, "Deep belief nets for natural language call-routing," in *Acoustics, Speech, and Signal Processing (ICASSP), 2011 IEEE International Conference on*, IEEE, 2011, pp. 5680-5683.
- [3] G. Mesnil et al., "Using recurrent neural networks for slot filling in spoken language understanding," *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, vol. 23, no. 3, pp. 530-539, 2015.

- [4] N. T. Vu, "Sequential Convolutional Neural Networks for Slot Filling in Spoken Language Understanding," in *INTERSPEECH*, 2016, pp.3250-3254.
- [5] S. Zhu and K. Yu, "Encoder-decoder with focus-mechanism for sequence labelling based spoken language understanding," in *Acoustics, Speech, and Signal Processing (ICASSP), 2017 IEEE International Conference on*, IEEE, 2017, pp. 5675-5679.
- [6] T. Luong, H. Pham, and C. D. Manning, "Effective Approaches to Attention-based Neural Machine Translation," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 1412-1421, Association for Computational Linguistics.
- [7] T. Hori et al., "Dialog State Tracking with Attention-Based Sequence-to-Sequence Learning," in *Proc. IEEE Workshop on Spoken Language Technology (SLT)*.
- [8] B. Liu and I. Lane, "Attention-Based Recurrent Neural Network Models for Joint Intent Detection and Slot Filling," in *INTERSPEECH*. 2016, pp. 685-689.
- [9] D. Bahdanau, K. Cho and Y. Bengio, "Neural Machine translation by jointly learning to align and translate," *CoRR*, abs/1409.0473.