Attention-based LSTM を用いた音声質問応答システムにおけるユーザーの質問意図理解 *

☆松好祐紀, 滝口哲也, 有木康雄 (神戸大)

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。使い始めの段階ではマニュアルを読むことになるが、それだけで十分使えるようになるのは困難であり、実際に動かしながら、使いながら慣れていくことが多いと思われる。この際に熟練者のアドバイス、 サポートを得ることが出来れば、習熟が早い。

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。

サポートは質問応答の形で行い、本稿では、ユーザーの意図を、質問から質問タイプ、質問キーワード分類を抽出することで推定する。推定には、Attention-base の LSTM を使用した。

実験として、質問文データ 487 文について、質問タイプ、質問キーワードの分類をアノテーションとして付与したものを学習、テストデータとした。学習データで学習を行い、テストデータに対して質問タイプと質問キーワード分類の推定率を調べた。また、質問タイプに関しては SVM、Random Forest でも同様の実験を行い、結果の比較を行った。

2　これまでの研究

我々のこれまでの研究では、システムをオセロゲームと質問応答システムで構成していた [1]。ユーザーがオセロゲームのプレイ中に質問が生じた場合、質問応答システムを起動し、質問をするという形であり、現在の研究も引き継ぎこの構成で行っている。

2.1　質問応答システム

これまでの研究では、質問応答システムは、ルールベースで構築しており、質問解析のために、ユーザーの質問の意図として質問タイプ、質問キーワードを定義していた [1]。

入力文からこれらをキーワードスポッティングの形で抽出し、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す。回答生成部では、キーワードとオセロプログラムからのパラメータを条件として、それらにマッチする回答テンプレートを選択し、そこにオセロプログラムのパラメータの値を埋め込む形で回答を生成する。

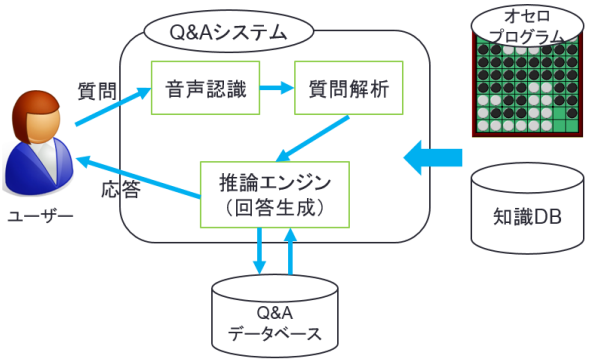

Fig. 1　提案システムの概略図

2.2　これまでの研究の問題点

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。

3　提案システム

3.1　システム概要

上記の問題を解決するためのシステムを Fig.1 に示す。ユーザーの質問が音声認識され、文字列の形で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。一度生成した回答はデータベースに格納しておき、同じような質問が来た際は、推論エンジンを用いずにデータベースから回答する。本稿では、質問解析部の研究について述べる。

3.2　質問解析部

質問の解析方法として、音声認識の結果より、質問文のフレーム、スロットを推定する。これまでの研究

* User's question intention understanding in voice question answering system using attention-based LSTM.
by MATSUYOSHI, Yûki, TAKIGUCHI, Tetsuya, ARIKI, Yasuo (Kobe University)

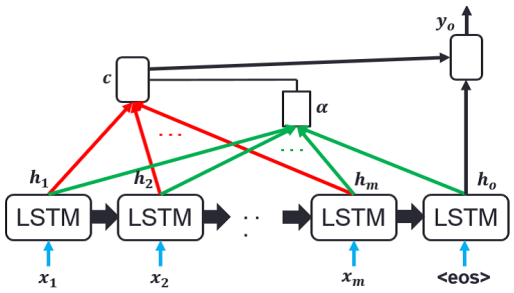

Fig. 2　Attention-based LSTM の概略図

に引き続き、フレームは質問の概要、ユーザーの大まかな意図を表すものとし、ユーザーの質問タイプと定義している。また、スロットは、質問中に出現するキーワード分類（以下、「質問キーワード分類」とする）と定義している。質問タイプに関しては、これまでの研究 [1] で用いたものを見直し、計 17 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対して生じる疑問に関する質問
- 場所: 盤面上の場所、座標に関する質問
- 結果: 指定した座標に打った場合の展開に関する質問
- 勝敗: 勝敗や形勢に関する質問

また、質問キーワード分類に関しても、従来研究のものを見直し、計 17 種類を定義した。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: b8、f7 など、オセロ盤面のマスの呼び方

質問タイプ、質問キーワード分類の推定には、Attention-based LSTM を用いた [2,3]。

3.3　質問タイプの推定

質問タイプ推定のモデルは Fig.2 のような、Attention-based LSTM である。まず、質問文の形態素をそれぞれを one-hot な単語ベクトルに変換する $(x_1 \dots x_m)$。そして、単語ベクトルが word embedding により分散表現に変換され、時系列的に LSTM に入力される。文末記号 <eos> が入力された時刻の中間層の出力を $\boldsymbol{h_o}$ とすると、質問文入力時に保持しておいた、以前の中間層の出力 $\boldsymbol{h_i}$(i=1,2,...,m) を利用して、Fig.2 の $\alpha(i)$ を計算する。

$$\alpha(i) = \frac{\exp(u_i)}{\sum_{j=1}^{m} \exp(u_j)} \tag{1}$$

$$u_i = \boldsymbol{W_1}^T tanh(\boldsymbol{W_2 h_i} + \boldsymbol{W_3 h_o}) \tag{2}$$

$\boldsymbol{W_1}$、$\boldsymbol{W_2}$、$\boldsymbol{W_3}$ は学習するパラメータである。$\alpha(i)$ を $\boldsymbol{h_i}$ の重みとし、コンテキストベクトル $\boldsymbol{c}$ を計算する。

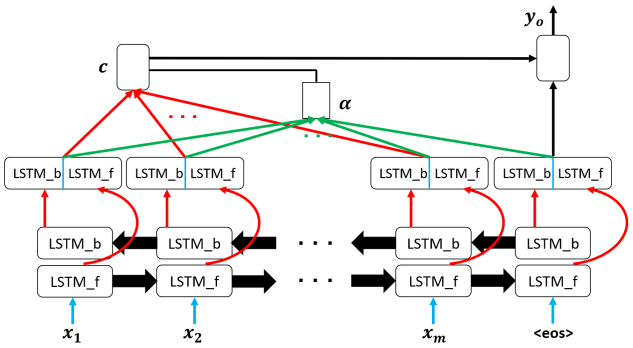

Fig. 3　Attention-based 双方向 LSTM の概略図

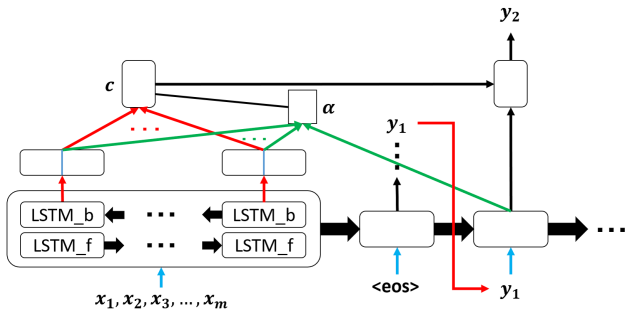

Fig. 4　質問キーワード分類推定モデルの概略図

$$\boldsymbol{c} = \sum_{i=1}^{m} \alpha(i) \boldsymbol{h_i} \tag{3}$$

$\boldsymbol{c}$ と $\boldsymbol{h_o}$ から、以下の式により $\widetilde{\boldsymbol{h}}_{\boldsymbol{o}}$ を計算する。

$$\widetilde{\boldsymbol{h}}_{\boldsymbol{o}} = tanh(\boldsymbol{W_4 c} + \boldsymbol{W_5 h_o}) \tag{4}$$

$\boldsymbol{W_4}$、$\boldsymbol{W_5}$ は学習するパラメータである。これに対して、線形作用素で重みを付けて、softmax を通したベクトルの中から、値が最大のものを質問タイプの推定値として得る。

$$y_o = argmax(softmax(\widetilde{\boldsymbol{h}}_{\boldsymbol{o}})) \tag{5}$$

また、Fig.3 のような、単方向 LSTM に時系列の逆方向の入力の情報も加える双方向 LSTM でもモデルを構築した。双方向 LSTM を導入した理由としては、双方向にすることで、Attention と同様に入力文の情報を大域的に取り扱えるからであり、Attention と組み合わせることで、さらにモデルの精度が良くなると考えたからである。順方向と逆方向の中間層の出力 $\boldsymbol{h_{fi}}$、$\boldsymbol{h_{bi}}$ を結合する際には、$\boldsymbol{W_6}$、$\boldsymbol{W_7}$ を学習パラメータとして、以下のように計算する。

$$\boldsymbol{h_i} = \boldsymbol{W_6 h_{fi}} + \boldsymbol{W_7 h_{bi}} (i = 1, 2, ..., m) \tag{6}$$

3.4　質問キーワード分類の推定

質問キーワード分類推定のモデルは Fig.4 のような、Attention-based LSTM Encoder-Decoder である [4,5]。Encoder 部分は、Attention-based 双方向 LSTM と同じであるが、Decoder 部分で <eos> が

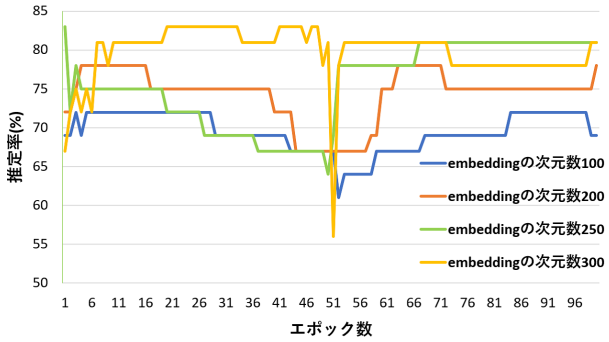

Fig. 5　word embedding の次元数を変化させることによる推定率の変化

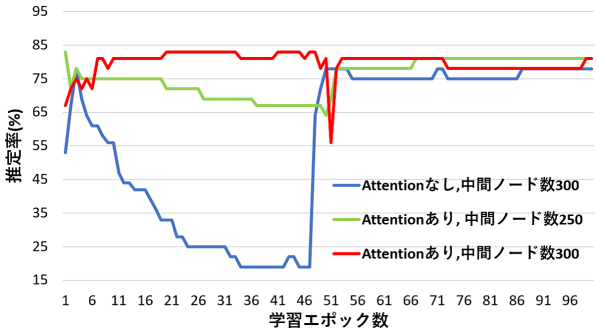

Fig. 6　Attention を導入したモデルと導入しないモデルでの推定率の比較

入力されると、質問キーワード分類が順番に出力されるように学習を行うモデルである。Attention は Decoder 部分の出力毎に計算され、推定に利用される。

4　実験

4.1　質問タイプの推定

学習データとしては、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実際に使用してもらった際に、ユーザーの質問を収集しているので、それらに対して質問タイプ、質問キーワード分類をアノテーションとして付与したものを用いた。ユーザーの質問文データは全部で 487 文あり、その内の 452 文を学習データ、残り 35 文をテストデータとし、それらに対応する質問タイプを教師データ、正解データとして用いた。

また、形態素解析には MeCab[6] を利用しているが、ユーザー辞書としてオセロの用語集を作成し、オセロ特有の単語や表現を取り扱えるようにしている。

学習したモデルにテストデータ 35 文を入力し、モデルから出力された質問タイプの推定値と正解データとが一致していた数を、テストデータの文の数 (35 文) で割った値に 100 を掛けた値をモデルの推定率 (単位は 100%) として実験を行った。

4.1.1　word embedding の次元数を変化させることによる推定率の変化

この実験では、word embedding の次元数を 100、200、250、300 と変化させて推定率の変化を調べた。これらのモデルは全て単方向 LSTM(Fig.2) である。学習はそれぞれ 100 エポック行い、1 エポック終了する毎にモデルを出力する。学習後に、出力した 100 個のモデル全てに対してテストを行った。

結果は Fig.5 のようになった。Fig.5 より、word embedding の次元数を増やすほど推定率が上がっていくが、次元数が 300 の場合では、少ないエポック数の時点ですでに推定率が高いモデルが学習されており、逆にエポック数を増やすと推定率が下がっていく。原因としては、学習データが少なく過学習が起きていることや、学習データの語彙数が 400 程度しかないことが挙げられる。

また、より推定率が高いモデルを学習出来ているのは、次元数が 300 の場合であり、83.3%となっているが、次元数が 250 の場合でも 80.6%の推定率のモデルが学習出来ており、あまり差はない。よって、word embedding の次元数は 250〜300 が妥当であると考えられる。

Table 1　他のモデルとの比較 (%)

	ルールベース	SVM	Random Forest
推定率	75.8	75.0	76.8

	LSTM	Attention-base LSTM
推定率	83.3	86.1

	双方向 LSTM	Attention-base 双方向 LSTM
推定率	88.9	88.9

4.1.2　Attention の導入による推定率の変化

この実験では、Attention を導入しない LSTM の中間層のノード数 300、Attention を導入した LSTM の中間層のノード数 250、300 の 3 つのモデルを比較した。実験方法は 4.1.1 と同様である。

結果は Fig.6 のようになった。Fig.6 より、Attention を導入しないモデルよりも、Attention を導入したモデルの方がより少ないエポック数で良い推定率のモデルを学習出来ていることが分かる。また、Attention を導入した LSTM の中間層のノード数 300 のモデルでは、10 エポック辺りからでも十分よい推定が出来るモデルが学習出来ており、最も良いモデルの推定率は 83.3%にまで向上した。

4.1.3　他のモデルとの比較

この実験では、Attention-based LSTM、Attention-based 双方向 LSTM の他にも、これまでの研究のルー

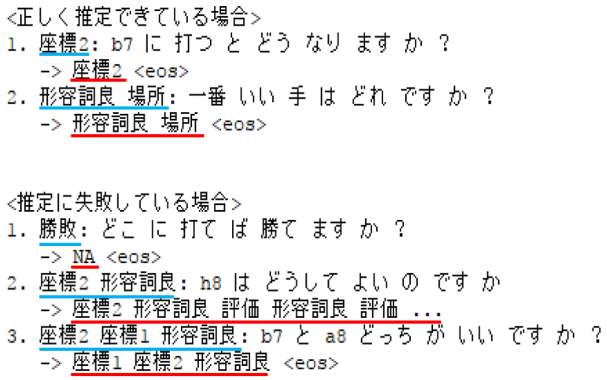

Fig. 7　質問キーワード分類推定の結果 (青下線が正解、赤下線が推定結果)

ルベース、Support Vector Machine(SVM)、Random Forest でも同様に質問文から質問タイプの推定を行い、比較した。SVM、Random Forest のハイパーパラメータに関しては、グリッドサーチを行い最も結果が良かったものを採用している。また、LSTM モデル全てにおいて、ドロップアウト (値は 0.5) を導入しており、前２つの実験結果よりも多少推定率が向上している。

実験の結果は Table 1 のようになった。ルールベース、SVM、Random Forest、Attention を導入しない LSTM で学習したモデルでは推定率が良くて 83.3%であったが、Attention を導入したモデルでは推定率が 86%を超えるモデルが学習出来るようになった。しかし、双方向 LSTM の場合に関しては、Attention を導入せずとも Attention を導入した双方向のモデルと同じ 88.9%の推定率のモデルが学習出来ており、Attention を導入しても推定率が向上しなかった。

4.2　質問キーワード分類の推定

質問キーワード分類に関しては、推定率の評価がまだ終わっていないので、テストデータを入力したことによってどのような結果が得られたかを述べる。実験は質問タイプの推定と同様に学習を 100 エポック行い、各エポック終了時にモデルを出力する。学習後に各モデルにテストデータを入力し、どのような結果が得られるかを調べる。

実験の結果、得られた出力は Fig.7 のようになった。上の２つが推定に成功した場合で、下の３つが推定に失敗した場合になる。推定する数が１つの場合はよく推定出来ているが、２つ以上になると推定率が急に落ちることが分析により分かった。２つ以上の推定の場合は、全く違う分類結果が出力されているものが多かったが、中には、Fig.7 の失敗した場合の２つ目、３つ目のように、似たようなキーワード分類で間違っている場合や、<eos> が上手く学習されておらず同じキーワード分類を繰り返す、といったような失敗も存在した。

5　おわりに

本稿では、ユーザーの質問意図として質問タイプと質問キーワード分類を推定するモデルの構築を行った。結果として、全体的に学習データ数が少なく、十分な学習が行えていないことが分かった。学習データ数を増やすことで、質問タイプ抽出に関してはより少ないエポック数でさらに良い推定率を有するモデルが学習でき、質問キーワード分類に関しては推定率が改善されると考えている。

質問タイプ推定に関しては、Attention を導入した結果、Attention を導入しないモデルより推定率の高いモデルが学習でき、Attention の有効性を示すことが出来た。しかし、想定したよりも推定率が向上せず、双方向 LSTM に関しては Attention を導入しない場合と同じであった。よって、実際に Attention がどのように影響しているのか調べるために、アテンションベクトルを解析する必要があると考えている。

質問キーワード分類の推定に関しては、推定率を向上させるために、Decoder 部分において一時刻前の推定結果のみを現時刻への入力としていたが、さらに一時刻前の情報まで入力するモデルの構築を試みる予定である。

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。

参考文献

[1] 松好祐紀, 滝口哲也, 有木康雄. ”ユーザー支援を目的とした音声質問応答システム 〜 オセロゲームの場合 〜,” 日本音響学会 2017 年春季研究発表会 2-P-11,2017-03
[2] Hori, T. *et al.* ”Dialog State Tracking with Attention-based Sequence-to-sequence Learning,” IEEE Workshop on Spoken Language Technology (SLT). December, 2016
[3] Minh-Thang Luong *et al.* ”Effective Approaches to Attention-based Neural Machine Translation,” arXiv preprint arXiv:1508.04025(2015)
[4] Bing Liu. Ian Lane. ”Attention-Based Recurrent Neural Network Models for Joint Intent Detection and Slot Filling,” Interspeech, 2016
[5] Dzmitry Dahdanau *et al.* ”NEURAL MACHINE TRANSLATION BY JOINTLY LEARNING TO ALIGN AND TRANSATE,” ICLR(2015)
[6] http://taku910.github.io/mecab/