

人の理解や習熟をサポートする音声質問応答システム

A voice question answering system supporting human understanding and proficiency

松好祐紀^{1*} 滝口哲也¹ 有木康雄¹
Matuyoshi Yūki¹ Takiguchi Tetsuya¹ Ariki Yasuo¹

¹ 神戸大学システム情報学研究科

¹ Graduate School of System Informatics, Kobe University, Department of Information Science

Abstract: The purpose of this study is to promote user understanding or skills by interactively responding to his questions uttered when using the system. In this paper, an interactive support system is especially considered for Othello games. In the previous study, the question answering system, constructed on a rule basis, has a limit in the ability to understand the user's intention. Therefore, in this paper, we propose "question type" estimation from the user's question by the Attention-based LSTM model. As a result, in the rule-based system, the recognition rate of the question type was 76%, while in the proposed model the recognition rate improved up to 86%.

1 はじめに

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る。

従来研究 [1] では質問応答システムをルールベースで構築していたが、ユーザーの意図を正確に把握するためには人手でルールを作成していくことに限界があることが分かった。そこで、本稿では Attention-based の系列的 LSTM モデルを用いて質問を解析し、ユーザーの意図を理解することを提案する。

2 提案システム

2.1 システム概要

システムの想定図を図 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、オセロのドメイン辞書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。

2.2 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、スロットを推定し、抽出する。本研究で

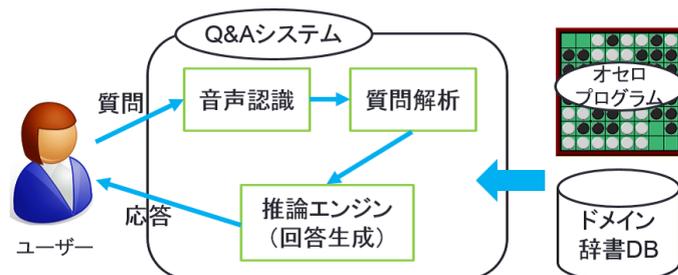


図 1: システム想定図

は、フレームはユーザーの大まかな意図を表すものであり、質問タイプとして定義している。質問タイプは、理由、定義、場所など 14 種類とした。また、スロットは、質問文中のキーワードの種類（「X 打ち」なら「オセロ用語」など）と定義している。本稿では、質問タイプの推定について述べる。

提案モデルは図 2 のような、LSTM を時系列的に並べたものに Attention を導入したモデルである。まず、質問文を形態素に分解し、それぞれを one-hot な単語ベクトルに変換する ($x_1 \dots x_m$)。単語ベクトルを時系列的に LSTM に入力する。文末記号 $\langle \text{eos} \rangle$ が入力された時刻において、LSTM の中間層の出力 h_o と、それ以前の間層の出力との関係を表すベクトル α を計算する。次に、 α から計算される全中間層の出力の情報を大域的に保持するコンテキストベクトル c を求めこれと LSTM の中間層の出力 h_o を基に、質問タイプの推定値 y を計算する [2,3]。

*連絡先：神戸大学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 自然科学総合研究棟 3 号館 805 号室
E-mail: matsuyoshi@me.cs.scitec.kobe-u.ac.jp

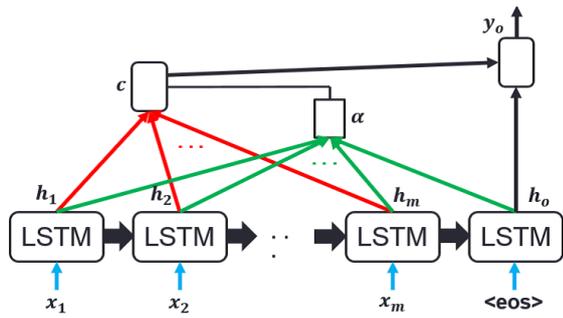


図 2: 提案モデルの概略図

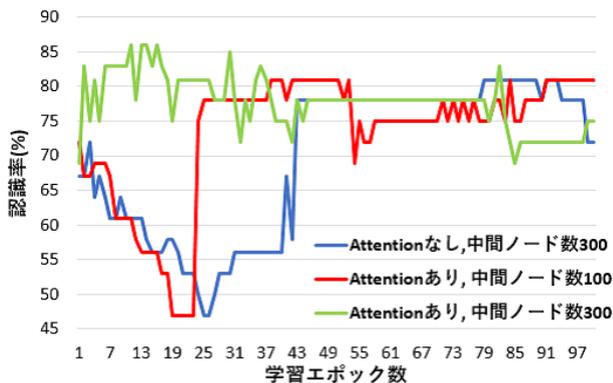


図 3: 実験結果

3 実験

3.1 実験手法

従来研究では、構築したルールベースのシステムを実際に使用してもらうことで評価したが、その際に収集した質問文毎に、質問タイプをラベル付けしたデータを本実験に用いた。質問文データは 487 文あり、その内 452 文を学習データ、残り 35 文をテストデータとした。実験としては、学習を 100 エポック行う際に、各エポック終了時にモデルを出力する。これらのモデルによる質問タイプの認識率を求めて、モデルの評価を行った。この実験を Attention なしの LSTM のノード数が 300 のモデルと、Attention ありの LSTM のノード数が 100、300 のモデルの 3 つのモデルに対して行い、比較した。

3.2 実験結果

実験結果は図 3 のようになった。図 3 より、Attention なしのモデルよりも Attention ありのモデルの方がより少ないエポック数で、良い認識率のモデルを学習出来ていることが分かる。また、Attention ありの LSTM のノード数 300 のモデルでは、数エポックで十分よく認識出来るモデルが学習出来ており、最も良いモデルの認識率は 86% にまで向上しており、Attention を導入することの有効性も示すことが出来た。

また、従来研究と比較した結果、ルールベースのシステムでの認識率が、76%であったことを考慮すると、今回比較した 3 つのモデルは、全てルールベースシステムよりも高い認識率のモデルが学習出来ている。

4 現在の取り組み

本研究では、フレームはユーザーの意図を表したものと考え、分類により質問タイプの推定を行った。この時、各質問タイプ毎に対応する質問キーワードも推定する必要がある。今回のモデルを展開して質問キーワードを推定することを考えたが、学習データのタグ付け量が多くなり困難である。そこで、学習データから自動で、タグ付けのための辞書データベースを作成する。手法は以下ようになる。

1. 学習データを形態素解析した結果から、python のライブラリ termextract を用いてオセロ用語の候補を抽出する。
2. その結果と「オセロ」で and 検索を行い、検索結果のスニペットにオセロ用語候補と一致する物があれば、「オセロ用語」を上位語としてその単語を辞書に登録する。
3. 形容詞やその他一般名詞は、形態素解析器の辞書を用いて分類し、辞書に格納する。

5 まとめ

本稿では、ユーザーからの質問を Attention-based の系列的 LSTM モデルを用いて解析し、質問タイプを推定するモデルの構築を行った。結果としては、学習データ数が十分でなかったが、ルールベースシステムでの結果を上回ることが出来た。今回は質問タイプを推定するモデルのみを構築したので、今後は質問のキーワードを推定するモデルの構築に取り組む。

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] Ravuri, Suman V., and Andreas Stolcke. "Recurrent neural network and LSTM models for lexical utterance classification." INTERSPEECH. 2015.
- [3] Minh-Thang Luong *et al.* "Effective Approaches to Attention-based Neural Machine Translation," arXiv preprint arXiv:1508.04025(2015)