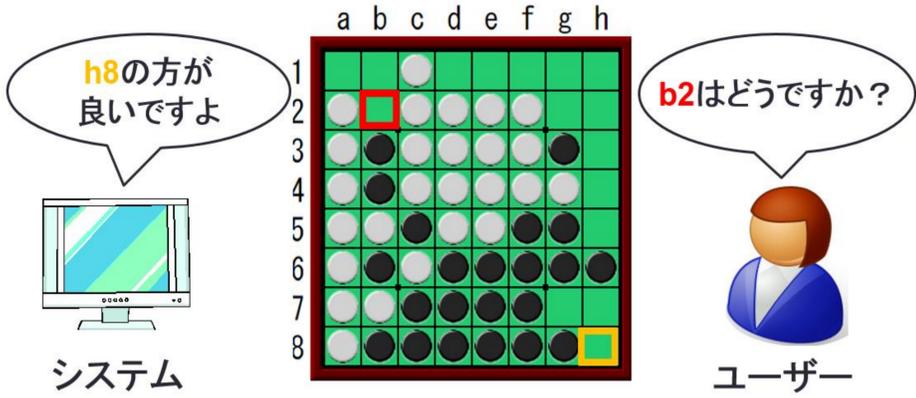


研究目的

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、ユーザーの理解を促進させる事が目的
 → 本研究ではシステムの具体例として、オセロに焦点を当てている

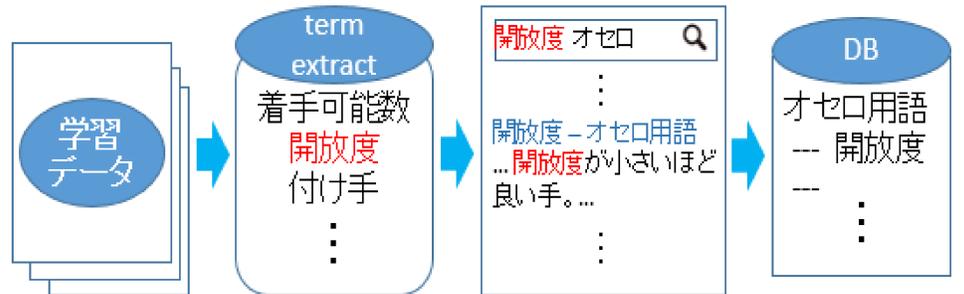


オセロをプレイする際、ユーザーの一手に対して対話的にアドバイス、サポートするシステム



- ・ スロットは、質問キーワードの種類(「X打ち」なら「オセロ用語」など)として定義している。
- ・ 質問タイプの推定に、Attention-basedの系列的LSTMモデルを用いた。
- ・ 時系列的な入力データの情報を大域的に保持するAttention機構を導入することで、質問タイプの認識精度を向上させることが出来ると考えた。

ドメイン辞書データベース



質問キーワード推定の際に必要なドメイン辞書データベースを、学習データから自動で作成することを考える。

- ・ 学習データを形態素解析した結果から、pythonのtermextractを用いてオセロ用語の候補を抽出。
- ・ オセロ用語候補と「オセロ」でand検索し、検索結果のスニペットにその候補と一致する単語があれば、「オセロ用語」を上位語として辞書に登録。
- ・ その他、質問キーワードとして必要な形容詞や一般名詞は、形態素解析器mecabの辞書を用いて分類し、辞書に登録。

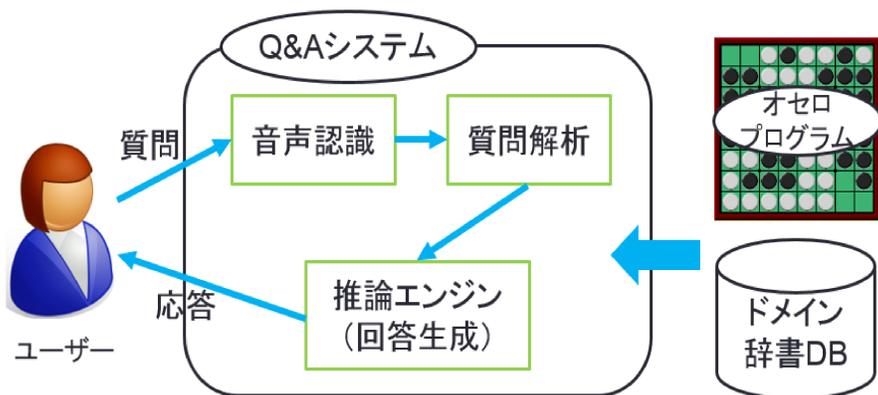
質問応答システム

今までの研究

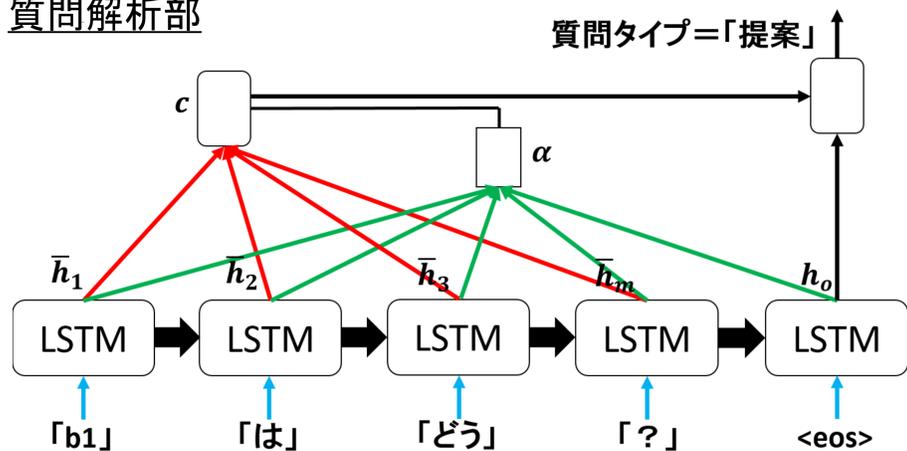
質問応答システムをルールベースで構築したが、以下のような問題点があった。

- ・ 極めて狭いドメインであるが、質問や回答に多様性があり、ルールベースで質問を解析、回答を生成するには限界がある。
- ・ オセロというドメインに固定されてしまう。

システム想定図



質問解析部



質問解析部では、質問文のフレーム、スロットを推定し、抽出する。

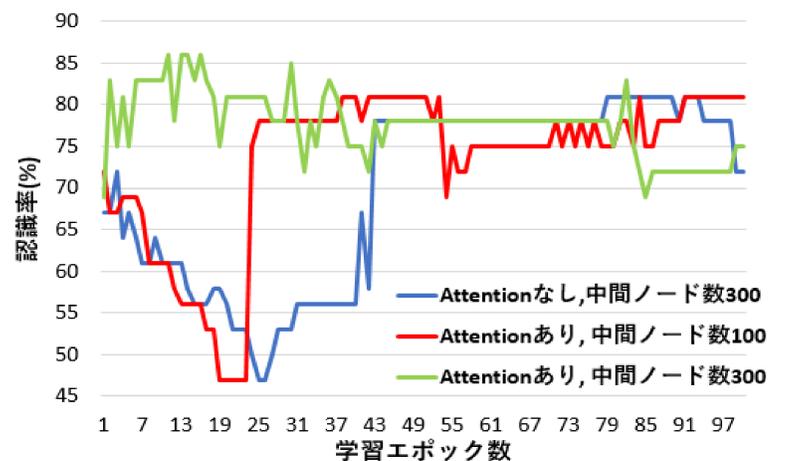
- ・ フレームはユーザーの大まかな意図を表すものであり、質問タイプとして定義している。質問タイプは、理由、定義、場所など 14 種類とした。

実験

実験概要

- ・ 従来研究でのシステムの評価の際に収集した、ユーザーの質問文487文に、質問タイプをラベル付けしたデータを使用。
- ・ 学習は100エポック行い、各エポック終了時に出力されたモデルの認識率を求めて、モデルを評価する。
- ・ 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。

実験結果



- ・ 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。
- ・ 従来のルールベースのシステムでは、認識率が76%であったが、Attentionなし, 中間層ノード数300のモデルでは86%に達した。