

# 遺伝的アルゴリズムと Word2Vec を用いた帰納的学習による 雑談システムの性能評価

黒田翔悟<sup>†</sup> 荒木健治<sup>††</sup>

<sup>†</sup> 北海道大学大学院情報科学院

<sup>††</sup> 北海道大学大学院情報科学研究院

## 1 はじめに

近年, 自然言語処理において雑談システムの研究が盛んに行われている. 特に, 深層学習を用いた雑談システムは Long Short Term Memory(LSTM) や Bidirectional Encoder Representations from Transformer(BERT) など, 様々なシステムが成果を上げている. 奥井らの研究 [1] では, LSTM を用いて対話を行っている個人に特化した応答アルゴリズムを提案し, その有効性を示している. しかし, 深層学習型システムの問題点としてモデルの中身がブラックボックス化してしまう点が挙げられる. そこで, 本稿では学習した中身をブラックボックス化せず, エラーの根本的な原因を人の目で見て理解できるアルゴリズムの構築を行い, その有効性を確認するために深層学習型の雑談対話システムとの比較実験を行った結果について述べる.

## 2 対話の実例からの学習

木村らの研究 [2] では, 対話の実例から単語単位で共通部分と差異部分を探索し, 応答ルールの生成を行っている. また, 荒木らの研究 [3] ではさらに性淘汰処理を導入し, 質の高いルールの生成を行った. そこで, 本稿では雑談対話コーパス [4] から文節単位の学習を行い, 変数部分に意味的特徴の付加を行うことで応答率と応答精度の向上を図った. また, 獲得されたルールにより応答できない場合は ELIZA を用いて応答を行う. 以下に各ルールの定義を示す.

### 2.1 共通ルール

対話例から文節の差異部分を変数化したものを共通ルールと呼ぶ. 変数部にはどのような文節が代入される可能性があるかを特徴づけるため, 文法的特徴と意味的特徴を記録する. 文法的特徴は以下の 2 つに分けられる.

1. 変数部自身の形態素
2. 変数部の係り先

意味的特徴は以下の 3 つに分けられる.

1. 変数自身の意味的特徴

parson, location, time タグを付与.

### 2. 変数同士の意味的特徴

差異部分の変数同士の類似度を算出し, 閾値以上となる文節を記録する. 類似度計算には Word2Vec を用い, 閾値は 0.7 以上とする.

### 3. 変数部と共通部内の各文節との関係性

変数部と共通部内の各文節の類似度を算出し, 閾値以上となる文節を記録する. 類似度計算には Word2Vec を用い, 閾値は 0.7 以上とする.

### 2.2 差異ルール

対話例の差異部分の文節の組み合わせを差異ルールと呼ぶ. 差異ルールでは入力文側の変数部に対し, その応答文の各文節を対応付け, 文法的特徴と意味的特徴を記録する. なお, 文法的特徴と意味的特徴の付与は共通ルールの変数部と同様の方法で行う.

### 2.3 帰納的学習

本手法によって獲得した共通ルール同士で学習を行い, より抽象化された共通ルール, 差異ルールを獲得する. その際, 新たに変数部となる文節は上記と同様の手法でルールを獲得する.

## 3 遺伝的アルゴリズムを用いたルールの生成

### 3.1 交叉処理

1 サイクルで得た対話と共通ルールを選択し, 一点交叉を行う. 一点交叉では各対話の文節を 1 つ選択し, 入れ替えることで新しいルールを獲得する. 文節の選択優先度は以下のように定義する.

1. 各ルールの文節のうち, 文法的特徴を満たし, Word2Vec による類似度が 0.7 以上の文節
2. 各ルールの文節のうち, Word2Vec による類似度が 0.7 以上の文節
3. 各ルールの文節のうち, 文法的特徴を満たす文節

### 3.2 突然変異

上記の優先度を無視して完全ランダムに交叉を行う処理を突然変異と呼ぶ. 突然変異では文法的制約や意味的制約を受けず交叉処理を行ってルールを獲得するため, 通常の交叉処理に比べ淘汰対象になる可能性が高い. したがって突然変異率は 5% とする.

3.3 淘汰処理

淘汰処理を行うために各ルールの適応度を以下のように定義する。適応度が75%未満のルールを淘汰対象とする。

$$\text{適応度} [\%] = \frac{\text{正応答回数}}{\text{正応答回数} + \text{誤応答回数}} \times a \times 100 [\%]$$

$$a = \begin{cases} 1 & (\text{対話例から獲得したルール}) \\ 1 - 0.8 \times \text{連続未使用回数} & (GA \text{ 処理により獲得したルール}) \end{cases}$$

4 比較対象

提案手法との比較実験を行うため、LSTMによる雑談対話システムを構築した。学習データには雑談対話コーパス [4] を使用した LSTM のモデル構築におけるパラメーターを以下に示す。

- ・単語ベクトル・隠れ層:1024次元
- ・損失関数:クロスエントロピー
- ・最適化:Adam

5 実験

5.1 実験設定

提案手法と LSTM による雑談対話システムを用いて印象評価実験を行った。

6名の評価者の入力に対し本手法と LSTM の両方を出力することで、どちらの出力が自然であったか1対話ごとに評価する。これを各評価者に対し20ターン行った。評価項目は、各出力文に対し評価者が

- 1.GAW-ILSD の応答の方が自然だった
- 2.LSTM の応答の方が自然だった
3. どちらの応答も自然だった
4. どちらの応答も自然でなかった

のいずれかに分類してもらうことで評価を行う。

5.2 実験結果

表1に実験結果を示す。なお、表中の「応答が自然と選択された数」は GAW-ILSD の場合は評価項目の1と3の総数、LSTM の場合は評価項目の2と3の総数を指す。

印象評価実験では評価者によってばらつきが見られた ( $\kappa = -0.203$ ) が、GAW-ILSD による応答ではベースラインとなる深層学習型システムを0.05ポイント上回った。なお、実験結果の有意差検定を行ったところ、 $p > 0.05$  より有意差は見られなかった。したがって、今後は変数部情報付加による応答性度の向上、大規模データの学習やアルゴリズムの文節から句単位への拡張によるルール生成数を増加などが今後の課題として挙げられる。

表1 :印象評価実験結果

	応答が自然と選択された数	選択率
GAW-ILSD	61/120	0.51
LSTM	55/120	0.46

5.3 エラー分析

印象評価実験により得られた結果から「GAW-ILSD の応答のほうが自然」とみなされなかった応答(以下、低印象応答と呼ぶ)についてエラー分析を行った。表2に低印象応答例と適用されたルールを、表3に適用ルール内の変数部情報を示す。

表2 :低印象応答例

入力	夏に山に行ったよ
出力	山が来たの?
適用共通ルール	@0 山に@1 山が@2
適用差異ルール	(夏に, 行ったよ) (来たの?)

表3 :@2の変数部情報

	形態素	係り先	タグ	類似度
@2	動詞 助詞	山が	NONE	NONE

低印象応答に適用されたルールを分析したところ、変数部分に情報が付加されていないルールが全体の約80%であった。これは、今回変数部分に使用した Word2Vec では網羅できない情報が学習の段階で欠損し、不適切な変数を代入してしまったことを示している。したがって、因果関係や文脈上のつながりなど、変数部分に付加する情報を増やすことで精度の向上が期待される。

参考文献

- [1] 奥井楓平, 中辻真:”LSTM を用いたパーソナル対話技術,” 人工知能学会全国大会 (第32回) 講演論文集, 4G2-02, 2018.
- [2] 木村泰知, 荒木健治, 桃内佳雄, 柘内香次:”遺伝的アルゴリズムを用いた帰納的学習による雑談を対象とした音声対話システム,” 電子情報通信学会論文誌 D-II, Vol. J84-D-II, No.9, pp.2079-2091(2001年9月).
- [3] Kenji Araki and Michitomo Kuroda :”Generality of Spoken Dialogue System Using SeGA-IL for Different Languages,” Proceeding of the IASTED International Conference COMPUTATIONAL INTELLIGENCE, pp.70-75, San Francisco, CA, U.S.A., Nov. 20-22, 2006.
- [4] 東中竜一郎, 船越孝太郎:”Project Next NLP 対話タスクにおける雑談対話データの収集と対話破綻アノテーション,” 人工知能学会 言語・音声理解と対話処理 (第72回) 研究会資料, pp.45-50, 2014-12-15.