

英語での応答問題を用いた文の接続の特徴調査と定量化

情報科学科 秋田 晃一郎

指導教員：山村 毅

1 はじめに

機械と人間との対話に関する研究は、自然言語処理の分野において古くから行われてきた。近年では特に自然な対話にその注目が集まっている。自然な対話とは、直感的には、「機械と感じさせない、まるで人間同士が行なっている」ような対話のことであるが、これを実現するためには、まず、文と文のつながり具合(文の接続性)を定量化する必要がある。本研究では、文の接続性の問題を「ある文の次に別の文が来る確率」という観点から捉え、TOEIC part2 の応答問題を対象に、質問文に対する適切な応答文を選ぶ手法の開発を行う。

2 応答問題

TOEIC part2 で出題される応答問題は質問文を聞いて選択肢の中から最も適切な応答文を選ぶ形式である。実際にはリスニング問題として出題されるので全て音声となっている。具体例を次に示す。

Q. Where is the meeting room?

(A) It's the first room on the right.

(B) To meet the new director.

(C) Yes, at two o'clock.

これは相手と英語で対話をする上で、相手が言ったことに対して意味の通った応答を選択する対話能力を見る問題である。また、一問一答の形式なので文同士の接続性が互いに直接反映され、本研究の対象として適していると考えられる。

3 提案手法

質問文を $q = (q_1, q_2, \dots, q_n)$ 、応答文を $r = (r_1, r_2, \dots, r_m)$ と表すとする (q_i, r_i はそれぞれ質問文、応答文の特徴を表す)。質問文に対する応答問題を選ぶという問題は質問文 q についての応答文 r の条件付き確率 $P(r|q)$ を考え、これを最大にするような応答文 (\hat{r} とする) を選ぶという問題に帰着される。 q_i と r_i がそれぞれの下で独立であると仮定すると、

$$\begin{aligned} \hat{r} &= \arg \max_r P(r|q) \\ &\approx \arg \max_{r_1, r_2, \dots, r_m} \prod_{i=1}^m P(r_i|q) \\ &= \arg \max_{r_1, r_2, \dots, r_m} \prod_{i=1}^m P(q|r_i)P(r_i) \\ &\approx \arg \max_{r_1, r_2, \dots, r_m} \prod_{i=1}^m P(r_i) \prod_{j=1}^n P(q_j|r_i) \end{aligned}$$

とすることができる。つまり、

$$f(q, r) = \arg \max_{r_1, r_2, \dots, r_m} \prod_{i=1}^m P(r_i) \prod_{j=1}^n P(q_j|r_i)$$

という評価関数を考え、これを最大にする応答文 \hat{r} を選べばよいことになる。

ここで「質問文に対する応答文」としては、応答文として正しいもの(正例)と応答文として正しくないもの(負例)の二つを考えることができるので、上記の評価関数 $f(q, r)$ として次の二種類を考えることができる。

1. 質問文 q に対する応答文としての r の正解度を表す評価関数 $f_p(q, r)$: 質問文に対する正しい応答文だけを用いて計算する。
2. 質問文 q に対する応答文としての r の不正解度を表す評価関数 $f_n(q, r)$: 質問文に対する誤った応答文だけを用いて計算する。

これらの評価関数を用いると正しい応答文を選ぶ問題には、次の三つのアプローチをすることができる。

1. 正解度の評価関数 $f_p(q, r)$ を最大にする r を選ぶ。(正例)
2. 不正解度の評価関数 $f_n(q, r)$ を最小にする r を選ぶ。(負例)
3. 正解度の評価関数 $f_p(q, r)$ と不正解度の評価関数 $f_n(q, r)$ の比 $\frac{f_p(q, r)}{f_n(q, r)}$ を最大にする r を選ぶ。(複合)

4 結果

1039 例(3 選択肢)の英文について、47 個の特徴を用いて応答文選択実験を行った。特徴は「where」、「in」、「on」、「when」、「at」などの単語の有無のほか情報利得の値により選定された単語の有無を用いている。評価には 10 分割検定を用いた。又、ゼロ頻度問題に対応するために加算スムージングを用いた。結果を表 1 に示す。

表 1 実験結果

学習データ	判定不能	正解	不正解	精度 (%)
正例	79	282	678	29.3
負例	37	410	529	40.9
複合	67	522	450	53.7

学習データに「正例」、「負例」、「複合」のいずれを用いた場合においても、選択肢に 3 つの応答文の評価値が同じになってしまう「判定不能」の場合があったため、精度評価はこれらを除いたものについて行なっている。

この表より、「正例」<「負例」<「複合」の順に精度がよくなっており、質問に対する応答が適切なものだけでなく適切でない文も評価に用いたほうが良いことがわかる。

5 まとめ

本研究では TOEIC part2 で出題される応答問題例文を用いて、特徴の抽出とその特徴を基にした定量化を行った接続性のあると考えられる特徴を抽出しナイーブベイズ分類器により実験を行った結果、学習として正しい応答文を使用したとき 29.3%、正しくない応答文を使用したとき 40.9%、そのどちらも使用したとき 53.7% の精度となった。今後の課題はデータ数を増やすことと構文解析や単語のカテゴリ分類により特徴を増やすことなどがある。