

## 207 対話処理における相互情報量の利用方法の一考察

木村 泰知† 荒木 健治‡ 桃内 佳雄† 棚内 香次‡  
 北学園大工† 北大工‡

## 1 はじめに

現在、多くの場面で自然言語による対話システムが望まれている。対話処理手法は多く存在するが、満足できる対話は行えていない。これは対話知識を全て計算機に予め与えることが難しいためである。そこで、生コーパスから対話に必要な情報の獲得を行う手法を提案する。人間の場合、問投詞などの重要度の低い単語を排除し、重要な単語を判断できる。このことを利用して、生コーパスから対応関係を獲得する際に相互情報量 [1] を利用することの有効性を考察する。本稿では、対話処理における相互情報量の利用方法を提案し、対話処理を表層的処理でどの程度行うことが可能かについて考察した。

## 2 相互情報量について

事象  $x$  及び  $y$  の共起確率がそれぞれ  $P(x), P(y)$  であり、 $x$  と  $y$  の共起確率が  $P(x, y)$  であるときに、 $x$  と  $y$  の相互情報量は (1) の式で定義される。

$$I(x, y) \equiv \log_2 \frac{P(x, y)}{P(x)P(y)} \quad (1)$$

この相互情報量を用いて、コーパス中の単語  $x$  及び  $y$  の出現を事象  $x$  及び  $y$  の生起とみなし、また、コーパス中で単語  $x$  と  $y$  が共起することを事象  $x$  と  $y$  の共起とみなす。(2) におけるように  $P(x), P(y)$ , 及び  $P(x, y)$  は、コーパス中での頻度  $f(x), f(y)$ , 及び  $f(x, y)$  をコーパスの総単語数  $N$  で割ったものを用いる。

$$P(x) = \frac{f(x)}{N} \quad P(y) = \frac{f(y)}{N} \quad P(x, y) = \frac{f(x, y)}{N} \quad (2)$$

コーパス中で単語  $x$  と  $y$  の共起については、いくつかの方法がある。A と B の対話で、A の入力「w1 w2 w3」に B が「w4」と応答した場合を考える。「単語数 2 の大きさの窓」の場合は、「w1 : w2」「w2 : w3」が共起となる。「対」で考える場合は、「w1 : w4」「w2 : w4」「w3 : w4」が共起となる。

## 3 実験

## 3.1 実験方法

実験データとして、RWCP の音声対話の書き下ろしコーパスを使用した。RWCP コーパスは、「車の購入」や「海外旅行の計画」などが含まれたコーパスである。RWCP コーパスを JUMAN[2] により単語に分割する。

この分割結果から、相互情報量を計算する。相互情報量は、「対」と「単語数 2 の大きさの窓」の 2 パターンで計算する。「対」の相互情報量は、応答に使用する単語を決定するために用いる。決定した単語に「単語数 2 の大きさの窓」の相互情報量が高い値の単語を連結し、応答文を生成する。被験者は大学院生一名とした。

## 3.2 結果・考察

被験者とシステムが 49 ターン対話を行なった。適切な応答はほとんど行なわれなかったが、対話の継続した部分も存在した。今回使用した RWCP コーパスは、人間と人間の模擬対話である。「人間と人間」と「人間とシステム」で対話に違いがあった。被験者がシステムに対して使用する単語数が少なかった。RWCP コーパスでは、平均単語数 21.7 単語 (括弧含む) であった。実験では、被験者の平均単語数は 4.2 単語であった。

RWCP コーパスは音声対話の書き下ろしであるため、コーパス自体が非文になっている部分も多く存在した。音声対話へ拡張を考えているため「RWCP コーパス」を使用した。非文の含まれていないコーパスとの比較も必要である。応答文を生成するための単語連結処理に「単語数 2 の大きさの窓」の相互情報量を利用した。対話としてはほとんど成立していなかったが、自然な単語連結が行われていた。単語連結する場合、「相互情報量」と「共起回数」を比較検討する必要がある。

単語の対応関係を獲得するために、「対」の相互情報量を利用した。適切な対応関係が獲得されたか判断が難しい。そのため、対応関係に絞った実験が必要である。

## 4 おわりに

相互情報量の利用方法はいくつも考えられる。今後、対話処理の有効な利用方法を確立する予定である。

## 謝辞

本研究の一部は、北海学園大学ハイテク・リサーチ・センターの研究費による補助のもとに行われた。

## 参考文献

- [1] 田中徳積 (監), "自然言語処理-基礎と応用-", コロナ社, 1999.
- [2] 黒橋 禎夫, 長尾 真, "日本語形態素解析システム JUMAN 使用説明書" version 3.61, 1999.