

## インターネット情報を利用した自由対話システムにおける応答文生成法

1M-02

黒柳智弘 石川 勉  
拓殖大学工学部情報工学科

## 1. はじめに

我々は、友達感覚の会話が楽しめる自由対話システムを目指して研究している[1]。現在のシステムは、すでに提案している概略推論法[2]を用い、入力文から獲得したユーザ情報をもとに、簡単な質問に答える。また、文の種類情報や吉凶情報を用いて気の利いたあいづちをうつ機能も実現している。しかし、現段階では、話題があらかじめ与えられた知識や獲得したユーザ情報に限定されてしまうという欠点がある。本報告ではインターネット情報を利用し、そこから獲得した情報と、入力文から抽出したキーワードをもとに、応答文を生成する方法を提案する。

## 2. 応答文生成法

図1に応答文生成の処理の流れを示す。本手法では、バックグラウンド処理として、インターネット上に流れている一定期間のニュース文書を保存しておく。この際、文書は政治、芸能、国際、経済、社会、スポーツの6ジャンルを対象とし、それぞれ独立したデータベースとして構築しておく。同時に、ジャンル毎の特徴単語リストを生成する。オンライン時には、入力文からキーワードを抽出し、特徴単語リストを利用して、ユーザがどのジャンルの話題をしているかを特定する。次に、特定したジャンルのデータベースから、キーワードを含む文書を抽出し、それをもとに応答文を生成する。

## 2.1 データベース構築

インターネット上から、前述した各ジャンルのニュース文書を獲得し、形態素解析した状態で保存する。常に最新の話題に対応可能とするため、保存する日数は7日間とし、キュー方式でデータを毎日更新する。

Generating Responses from the Internet News for a Dialogue System  
Tomohiro Kuroyanagi, Tsutomu Ishikawa  
Department of Computer Science, Takushoku University

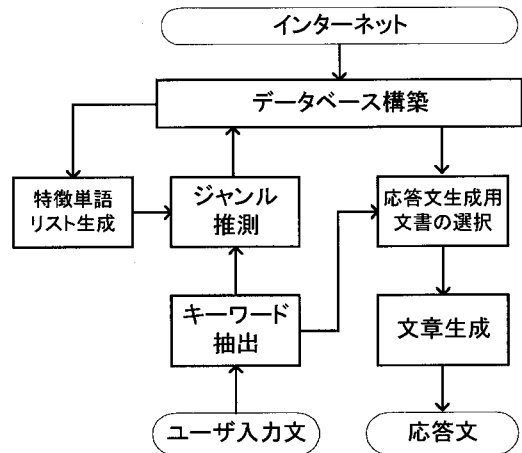


図1. インターネット情報からの応答文生成

## 2.2 特徴単語リスト

ジャンル推測を行うため、特徴単語リストと呼ぶ照合用ファイルを作成する。このリストでは、tf-idf 法[3]を用いて、各々のジャンルのデータベースから抽出した単語に重要度を持たせておく。なお、リストは一般単語(動詞、一般名詞)、固有名詞のそれぞれについて作成する。

## 2.3 ジャンル推測

入力文から一般単語、固有名詞を抽出し、これらをキーワードとして特徴単語リストと照合する。この際、固有名詞は単純に照合するが、一般単語に関しては、単語間の意味の類似性を判別するツールである概念ベース[4]を利用し、単語の類似度計算を行い、この結果を利用する。

具体的には、まず下式のように各キーワードについて、ジャンル毎に特徴単語リストの重要度と類似度の積の最大値を求め、そのキーワードの各ジャンルに対する加算得点とする。この得点をジャンル毎に、すべてのキーワードについて総計し、その得点が最大値となったジャンルを、ユーザが話題にしているジャンルと特定する。

$$J_a = \max(P_i \times D_i) \quad (1)$$

( $J_a$ :ジャンルの加算得点,  $P_i$ :特徴リスト  $i$  番目の重要

度, Di:特徴リスト i 番目の単語とキーワードの間の類似度)

## 2.4 応答文生成用文書の選択

入力文にマッチした応答文を生成するため, ジャンル毎のデータベースから, 最適な文書を選択する。

具体的には, まず入力文から抽出した固有名詞のキーワードをメインキーワード, 普通名詞と動詞のキーワードをサブキーワードとし, 特定したジャンルのデータベースにある索引ファイルを調べ, メインキーワードを含む全てのタイトルを抽出する。この索引ファイルにはタイトルと文書のファイル名が記載されている。次に, 抽出されたタイトルの中の単語とサブキーワードの間の類似度を総計し, この得点が最も高かった文書を応答文生成用の文書とする。

## 2.5 文章生成

入力文のメインキーワードを主語とする。また, 2.4で選択した文書からそのキーワードに対する動詞, 目的語を抽出する。それらを連結させることにより文章を生成する。以下, これらの処理について述べる。

### (1) 単語の文節化と文章の組み立て

2.1で行った形態素解析は, 意味のある最小単位の単語に文章を分解する。したがって, これらの文章を所有格“の”以外の助詞, 句読点等の記号を区切りとし, 文節にまとめる。

次に, 目的語を含む文節と, 動詞を含む文節を主語に連結することにより文章を組み立てる(図2)。

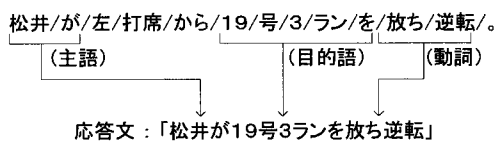


図2. 文節化と文章の組み立て

### (2) 主語が欠落した文書

主語が欠落した文書も文章生成の対象とする。第一文に主語があり, 第二文に主格助詞(は, が, も)がなければ, その第二文も第一文と同じ主語とみなし, 前述した方法で文章生成を行う。

### (3) 自然な語句の付加

応答文を柔らかい文章表現にするため, 生成した文章の語尾に注目し, 適切な語句をその語尾に追加する。例えば, 生成した文章の語尾が用言性名詞の場合(例“逆転”), “したよ”という語句を追加する。

## 3. 実験

実験システムを構築し, 本手法の動作確認を行った。完璧ではないが, ある程度入力文に関連する応答文が生成できた。以下, 生成された応答文の例を示す。

### 成功例)

入力文: 「小野は試合に出て活躍しているみたいだね。」  
 応答文: 「小野は堅守に徹底しました。」

### 失敗例)

入力文: 「浜崎あゆみはいつ新曲をだすのかなあ。」  
 応答文: 「浜崎が制覇したよ。」

入力文にマッチしない応答文が生成される場合の多くは, 本システムが複雑な構文解析や意味解析を行っていないため, 文章の組み立て段階で失敗した場合である。今後はこれらについて改善していく予定である。

## 4. 結び

本論文では, 自由対話システムにおいて, インターネット情報を利用し, 入力文にマッチした応答文を生成する手法を提案した。また, 実験でその基本動作を確認した。

### 参考文献

- [1] 小林, Nguyen V. H, 塩野, 石川: 感情表現と質問応答機能を備えた知的自由対話システム, 人工知能学会研究会資料, SIG-SLUD-A003-1, pp. 1-8, 2001.
- [2] Nguyen V. H, 石川勉, 阿部明典: 知識の類似性を利用した概略推論法, 電子情報通信学会論文誌 DI, Vol. J84-D-1, No. 4, pp. 389-400(2001).
- [3] Salton G. and Allen J.: Text Retrieval Using the Vector Processing Model, 3rd Annual Symposium on Document Analysis and Information Retrieval, pp. 9-22(1994).
- [4] 笠原, 松沢, 石川: 国語辞書を利用した日常の類似性判別, 情処学会論文誌, Vol. 38, No7, pp. 1272-1283, 1997.