

# Weblog を用いた対話的リアクションシステムの構築

伴昌彦<sup>†</sup>, 吉田文彦<sup>‡</sup>, 内田理<sup>†</sup>, 菊池浩明<sup>†</sup>, 中西祥八郎<sup>†</sup>

<sup>†</sup> 東海大学電子情報学部, <sup>‡</sup> 東海大学文学部

e-mail : masa@kuwa.ep.u-tokai.ac.jp

## 1 はじめに

近年、対話システムの研究[1]~[5]が盛んに行われている。その中で[3], [4], [5]はユーザとの自由な対話を対象とした対話システムであるが、[3], [4]は時事問題などの話題に対応ができない、[5]は「やめましようのですね」のように応答文の表現が不自然になることがあるという問題がある。

そこで、本稿では Weblog を用いた対話的リアクションシステムを提案する。2005 年 1 月の時点で少なくとも 60 万以上あるといわれている Weblog には多くの事柄に関する様々な意見や論評などが書かれている。このため、Weblog からシステムの応答文を求めることにより、時事問題をはじめとした多様な話題への対応が可能となる。また、Weblog は人によって書かれているため応答文の表現が不自然になるということもない。

なお、本システムでは対話内容の解析を行っていない。そのため、入力文と応答文との関係は「対話」よりも、何か反応するという意味の「リアクション」の方が適当であるので、「リアクションシステム」とした。

## 2 リアクションシステム

### 2.1 システム概要

本システムの処理プロセスを図 1 に示す。

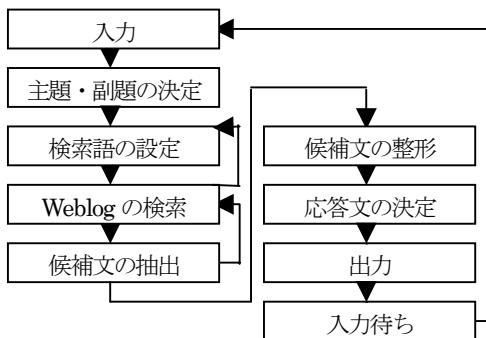


図 1 システムの処理プロセス

まず、主題・副題決定プロセスを用いてユーザの入力文の主題・副題を決定する。次に主題・副題から検索語設定プロセスを用いて Web 上の Weblog の検索に用い検索語を設定する。次に検索語を用いて Web 上の

Weblog を検索する。さらに検索された Weblog に付いているタイトルから Weblog 検索プロセスを用いて応答文の候補文を抽出する Weblog を決定する。次に決定した Weblog の本文から候補文抽出プロセスを用いて応答文の候補文を抽出する。さらに、候補文を応答文として相応しい形に整形する。最後に応答文選択プロセスを用いて応答文を決定して出力する。その後、再びユーザから入力があれば、これらの処理を繰り返す。

### 2.2 主題・副題決定

#### 2.2.1 入力文(最初)の主題・副題決定

本システムでは Weblog から応答文を求めするため、Web 上の Weblog を検索する語が必要となる。さらに、検索語は入力文の内容が含まれている語が適当である。本稿ではその文の内容が含まれている語を「主題」とし、主題の補助的な語を「副題」とする。主題・副題になる語はその文中の複合名詞及び「名詞一般、名詞-サ変接続、名詞-固有名詞」の品詞の語とし、これらの語をまとめて「主副名詞」とする。さらに、主題・副題を決定する文を「対象文」とする。なお、品詞は茶筌[8]の品詞体系を用いた。ここで、以下に示す主題・副題決定プロセスはその対話における最初の入力文の主題・副題を決定するものである。

- 1) 入力文が複数(「。」が複数)ならば、最後の入力文を「対象文」とする。
- 2) 対象文に主副名詞がない場合は主題・副題はなしとする。
- 3) 対象文に主副名詞があり、格助詞「の」がない場合は全ての主副名詞を主題とする。
- 4) 対象文に主副名詞があり、格助詞「の」があるが主副名詞と繋がっていない場合は全ての主副名詞を主題とする。
- 5) 対象文に主副名詞があり、それに格助詞「の」が付く場合は「の」の後に付く主副名詞を主題、前に付く主副名詞を副題とする。この時、「の」の前に付く語が主副名詞ではない場合は、「の」の後に付く語を主題とする。また、「の」の前に付く語が主副名詞ではない場合は、主題のみとなる。

なお、対象文に格助詞「の」が複数ある場合は最後の格助詞「の」から主題・副題を決定する。

- 6) 4)の場合において、主題と「他の主副名詞」の間に格助詞「と」または「や」が付く場合は「他の主副名詞」も主題とする。同様に副題の場合は「他の主副名詞」も副題とする。

### 2.2.2 二番目以降の入力文の主題・副題決定

二番目以降の入力文の主題・副題の決定に2.2.1のプロセスは用いないことにする。それは、二番目以降の入力文にはその直前の応答文の内容が含まれていることが多いからである。例えばシステムの応答文「春だね。」に対し、ユーザの入力文が「桜が咲くね。」とする。この「桜」は「春」と関連している語であるため、ユーザの入力文に出てきた語と考えられる。さらに、「春」と「桜」はそれぞれの文の主題である。このようにその時点でのシステムの最後の応答文の主題・副題とそれに対するユーザの入力文の主題・副題に関連がある場合がある。この場合は2.2.1とは別の主題・副題決定プロセスを用いてユーザの入力文の主題・副題を決定する必要がある。以下に関連による主題・副題決定プロセスを示す。

- 1) 対象文の主題・副題の「候補」を2.2.1の主題・副題決定プロセスを用いて決定する。
- 2) 対象文の主題・副題の候補とその時点でのシステムの最後の応答文の主題・副題との関連を調べる。
- 3) 関連があれば、対象文の主題・副題の候補とその時点でのシステムの最後の応答文の主題・副題で関連しているそれぞれの語を残す。
- 4) 対象文の主題・副題の候補の中で残ったものをその対象文の主題として決定し、その時点でのシステムの最後の応答文の主題・副題の中で残ったものを対象文の副題として決定する。
- 5) 関連がなければ、1)で決定した主題・副題の候補をそのまま対象文の主題・副題として決定する。

### 2.3 検索語の設定

検索語は2.2で決定した主題・副題から優先順位をつけて設定する。以下にそのプロセスを

- ①「主題が一語のみで副題がない場合」
- ②「主題が複数あり、副題がない場合」
- ③「主題・副題が一語以上ある場合」
- ④「主題・副題がない場合」

の4つの場合で示す。

#### ①の場合

この主題を優先順位1の検索語とする。優先順位2の検索語はない。

#### ②の場合

- 1) 全ての主題を優先順位1の検索語とする。
- 2) 主題の中で入力文の最後にある主題を優先順位2の検索語とする。
- 3) 入力文中で2)に該当する主題の一つ前の主題を優先順位3の検索語とする。
- 4) さらに、入力文中でその一つ前に主題がある限り、順に優先順位をつけて検索語とする。

#### ③の場合

- 1) 全ての主題・副題を優先順位1の検索語とする。
- 2) 優先順位2からは主題が一語ならば①、複数ならば②と同様のプロセスで主題から検索語と優先順位を設定する。
- 3) 2)の後の優先順位として、副題が一語ならば①、複数ならば②のプロセスで『主題』を『副題』に置き換えて副題から検索語と優先順位を設定する。

#### ④の場合

検索語を設定できないため、Web上のWeblogから応答文を抽出することはできない。よって、その入力文に対する応答文は2.6.2のプロセスを用いて決定する。

なお、ここで設定した検索語の順位を $S(i)$ とする。

## 2.4 応答文の候補文の抽出

### 2.4.1 候補文を抽出するWeblogの検索

2.3で設定した $S(i)$ の検索語とWeblog専用の検索エンジンを用いてWeb上のWeblogを検索する。本システムではWeblog専用の検索エンジンとして「未来検索ライブラリ」を使用した。さらに、検索されたWeblogに付いているタイトルからWeblog検索プロセスを用いて候補文の抽出に用いるWeblogを決定する。以下にそのプロセスを示す。

- 1) 検索結果の上位 $\alpha$ 件までのWeblogを候補文の抽出対象とする。本システムでは $\alpha=500$ と設定した。
- 2) Weblogのタイトルに含まれていればよい語とその優先順位 $W(j)$ を設定する。この設定には2.3の検索語設定プロセスを用いる。つまり、 $W(j)=S(i)$ 、 $(j=i)$ である。
- 3)  $S(1)$ の語でWeb上のWeblogを検索する。Weblogが検索されなければその都度 $i=i+1$ とし、 $S(i)$ の語でWeb上のWeblogを検索する。 $i$ を最大にしてもWeblogが検索されなければその入力文に対する応答文は2.6.2のプロセスを用いて決定する。
- 4) 3)で検索された最大で $\alpha$ 件のWeblogから $W(1)$

の語が含まれている Weblog のタイトルを全て見つける。見つからない場合はその都度  $j=j+1$  とし、 $W(j)$  の語が含まれている Weblog のタイトルを見つめる。  $j$  を最大にしても  $W(j)$  の語が含まれている Weblog のタイトルが見つからなければ  $i=i+1$  とし、2.4.1 のプロセス 3) に戻る。ここで、 $i$  を最大にしても  $W(j)$  の語が含まれている Weblog のタイトルが見つからなければその入力文に対する応答文は 2.6.2 のプロセスを用いて決定する。

5) 見つかった場合はさらにその中で以下の全てのルールに該当するタイトルの Weblog を全て候補文の抽出に用いる Weblog に決定する。

- ・タイトル中で  $W(j)$  の語が複合名詞の一部ではない。
- ・タイトル中で  $W(j)$  の語が  $W(j)$  の語以外の語と格助詞「の」で繋がっていない。
- ・タイトル中で  $W(j)$  の語が括弧等の記号で括られていない。

#### 2.4.2 候補文抽出

2.4.1 で決定した候補文の抽出に用いる Weblog の本文から候補文抽出プロセスを用いて候補文を抽出する。ここで、候補文を Weblog の二文目や三文目から抽出するとその文には前の文の内容が含まれていることが多いため応答文には適さない。そのため、前の文の内容が含まれない一文目から候補文を抽出する。さらに、Weblog においても空行は内容が変わるときに存在するものと判断したため、空行後の一文目からも候補文を抽出する。以下にそのプロセスを示す。

- 1) 候補文に含まれていればよい語とその優先順位  $C(k)$  を設定する。この設定には 2.3 の検索語設定プロセスを用いる。つまり、 $C(k)=S(i)$ , ( $k=i$ ) である。
- 2) 2.4.1 で決定した全ての Weblog の一文目と空行後の一文目から  $C(1)$  の語が含まれている文を見つめる。見つからない場合はその都度  $k=k+1$  とし、 $C(k)$  の語が含まれている文を見つめる。  $k$  を最大にしても  $C(k)$  の語が含まれている文が見つからなければ、 $j=j+1$  とし、2.4.1 のプロセス 4) に戻る。ここで、 $j$  を最大にしても  $C(k)$  の語が含まれている文が見つからなければ  $i=i+1$  とし、2.4.1 のプロセス 3) に戻る。ここで、 $i$  を最大にしても、 $C(k)$  の語が含まれている文が見つからなければその入力文に対する応答文は 2.6.2 のプロセスを用いて決定する。
- 3) 見つかった場合はさらにその中で以下の全ての

ルールに該当する文を全て候補文として抽出する。

- ・その文に接続詞[7]、指示語がない
- ・その文に「 」や『 』がない
- ・その文で  $C(k)$  が複合名詞の一部ではない
- ・その文に動作を表す動詞(歩く、食べる)がない
- ・その文に過去形の表現がない
- ・その文の文末に助詞「と」「に」「へ」がない
- ・その文で  $C(k)$  の語が  $C(k)$  の語以外と格助詞「の」で繋がっていない

#### 2.5 語尾変換による文の整形

Weblog は様々な人達が自由に書いているため文体の「ですます体」と「である体」は統一されていない。このため、Weblog から抽出した候補文をそのまま応答文にしてしまうと、同じ対話において「ですます体」の文と「である体」の文が混在してしまい不自然な対話になる。そこで、本システムでは候補文の文体が「ですます体」の場合は「だ体」、「である体」または「言い切りの形」に変換する。例えば「林檎でしょう。」は「林檎だろう。」に、「食べます。」は「食べる。」に変換する。なお、変換に用いる語尾変換システムに関しては本稿の主旨からそれるため、紙面の都合上割愛する。

#### 2.6 応答文の決定

##### 2.6.1 候補文からの応答文の決定

2.5 で整形した候補文の中から応答文選択プロセスを用いて応答文を決定する。人間同士の対話においても、互いに考えや感想を述べ合うとその対話は長く続くものになる。本システムでは「思う、考える、感じる」などの内的動作動詞が含まれる候補文を考えや感想を述べている文とする。さらに、「桜は美しいね。」などのように、その事柄についての感想を形容詞を用いて表現することがあるため、形容詞が含まれる候補文も感想を述べている文とする。以下に応答文決定プロセスを示す。

- 1) 内的動作動詞(思う、感じる、考える)がある候補文があればその文を応答文に決定する。
- 2) 1)に該当する候補文がない場合は形容詞がある候補文を応答文に決定する。
- 3) 2)に該当する候補文がない場合は候補文の中で文字数が一番少ない候補文を応答文に決定する。
- 4) 1), 2), 3)において該当する候補文が複数ある場合はその中で文字数が一番少ない候補文を応答文に決定し、それも複数ある場合は抽出した順が早い方の候補文を応答文に決定する。

##### 2.6.2 その他の応答の決定

Weblog から応答文を抽出できない場合は「肯定応答

文」または「相槌応答文」で応答をする。以下にそのプロセスを示す。

- 1) 入力文の語尾が終助詞「～ね、～ねえ、～ねえ、～なあ、なあ、～よね、～よな」である場合は「そうだね」「そうね」といった入力文の内容を肯定する肯定応答をする。
- 2) 1)に該当しない場合は単に「うん、うん。」「へえ～。」といったユーザにさらに対話を促す相槌応答をする。

### 3 評価と考察

本システムで作成した対話内容を以下に示す。

**User:**「映画を見に行って寝てしまった・・・」  
**System:**『映画ってほんとうにいい。』  
**User:**「そうね。でも、つまらない映画ほど眠くなる。」  
**System:**『まあ、たまには映画なんてのも悪くないものだ。』  
**User:**「その後、ボーリングに行った。」  
**System:**『ボーリング好きなんだけど・・・』  
**User:**「あ、なんか、雨降ってきた。」  
**System:**『夜、雨。』  
**User:**「まさか、台風・・・」  
**System:**『なんか・・・ほんとに台風みたいだ。』  
**User:**「マジか・・・!?!」  
**System:**『意外と、共感をえるコトが多いのが雨がクライじゃないとゆうことだ。』

このような対話A～Eを5つ作成し、10人にA～Eの

表1 評価結果

	A	B	C	D	E
①	2.5	3.0	2.8	3.5	2.3
②	2.3	3.2	3.8	3.5	3.0
③	3.5	2.5	3.0	2.2	3.7
④	3.3	2.3	2.7	1.3	3.5
⑤	3.0	3.0	2.5	2.7	2.7
⑥	3.0	2.7	2.5	3.0	2.7
⑦	3.7	2.2	2.2	2.2	1.0
⑧	2.7	2.2	3.2	2.8	2.3
⑨	2.5	2.5	1.8	2.3	1.5
⑩	2.8	2.7	3.5	4.0	2.7
平均	<b>2.9</b>	<b>2.6</b>	<b>2.8</b>	<b>2.8</b>	<b>2.5</b>

対話中に6組ある「入力文に対する応答文」の内容を次の五段階で評価してもらった。

『1:ダメ 2:どちらともいえない 3:まあまあ  
 4:よい 5:とてもよい』

その結果を表1に示す。但し、A～Eは作成した対話、①～⑩はアンケートの回答者である。さらに、表の値は対話A～Eにおける6組の「入力文に対する応答文」の内容の評価の平均値である。

ここで、図2の「入力文に対する応答文」の評価を④の回答者は『5』、⑦の回答者は『1』としており、評価には個人差があることが分かる。

**User:**「そうね。でも、つまらない映画ほど眠くなる。」  
**System:**『まあ、たまには映画なんてのも悪くないものだ。』

図2 「入力文に対する応答文」例

それが、全体的に評価の低い原因の一つであると考えられる。しかし、談話解析といった処理を一切施さないシステムの応答文の内容の評価としては全体的に『3:まあまあ』に近い値なので、Weblogは応答文を求める対象としては使用可能であるといえる。

### 4 おわりに

システムの応答文をWeblogから求めるリアクションシステムの提案した。本システムでは対話文の内容を解析する処理を行っていない。そのため、今後の課題は、このシステムに談話解析等の処理を加え、人間同士のよ様な対話が行えるようにすることである。

### 参考文献

- [1] 北岡,加藤他/車載機器用音声対話システム,情報処理学会研究報告. SLP, 音声言語情報処理, Vol. 96 Num. 74 pp.57-62 (1996.07)
- [2] 伊藤,酒井他/全文検索による CD-ROM ガイドブックの内容検索,情報処理学会研究報告. 情報学基礎研究会報告, Vol. 93 Num. 67 pp.1-8 (1993.07)
- [3] 江部,乾他/会話データとのキーワードマッチングを行い、応答文を決定する対話システム,情報処理学会第 58 回全国大会公演論文集(2),pp.281-282 (1999)
- [4] 江部,乾他/発話の働きとキーワードから応答生成を行う事例ベース自由対話系,情報処理学会研究報告. FI, 情報学基礎, Vol. 2001 Num. 20 pp.23-30 (2001.03)
- [5] 木村,荒木他/遺伝的アルゴリズムを用いた帰納的学習による音声対話システムの有効性について,電子情報通信学会技術研究報告. TL, 思考と言語, Vol. 99 Num. 691 pp.41-48 (2000.03)
- [6] 長尾真:自然言語処理,岩波書店,(1996)
- [7] 森岡健二他:「接続詞と感動詞」,明治書院,(1973)
- [8] 奈良先端科学技術大学院大学松本研究室:「日本語形態素解析システム 茶釜」  
<http://chasen.aist-nara.ac.jp/hiki/ChaSen/>