

## 相手の嗜好にあった話題を提供する自動発話システムの開発

林 輝大† 奥村 紀之†

†長野工業高等専門学校 電子情報工学科

## 1 はじめに

本研究は、人間と計算機とがより自然な会話を行えるようなシステムを開発することを目的としている。

計算機が話題を提供するための手法として、「発話のサイクル」と「背景的知识」を提案し、発話のサイクルにおける話題決定の評価を行う。

## 2 提案手法

既知の会話ロボット [1] は受動的なシステムが一般的である。本研究では計算機が能動的に話題を提供可能なシステムとして、「発話のサイクル」、「背景的知识」を提案する。「発話のサイクル」とは図 1 に示すような循環するモデルと定義する。

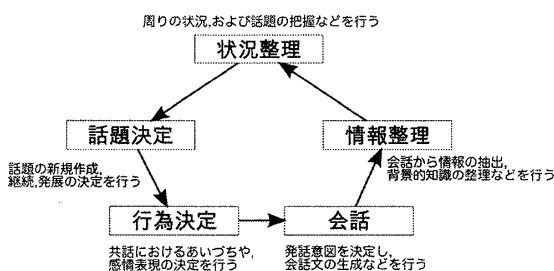


図 1: 提案する発話のサイクル

## 2.1 背景的知识

人間は会話の際、知識や記憶として頭の中にある膨大な情報から、周りの状況に合わせて会話を展開する。その膨大な知識を本研究では「背景的知识」と呼ぶ。

本研究では「一般知識」、「記憶」、「自他の個人データ」の 3 つに大別し、定義している。

## 2.2 話題決定

図 1 における「話題決定」の主な役割は、会話をしているある時点での話題に対し、次に提供すべき話題を背景的知识を適宜参照し決定することである。

話題決定は 3 つの手順で構成され、順に「候補集合の取得」、「話題語群抽出」、「候補の選定」である。以

降、それぞれについて提案する手法を述べる。

## 2.2.1 候補集合

話題語を抽出するための候補集合は、2 種類の語の集合から構成される。1 つ目は背景的知识に蓄えられた会話相手の情報であり、現在の話題語に関連している語の集合である。2 つ目は大規模概念ベース [3] から現在の話題語に対する二次属性を取得した語の集合である。

## 2.2.2 抽出方法

候補集合から話題語として適切な語を抽出するために優先度をつける。その優先度を算出する 4 つの手法を提案する。なお、優先度は 0~1 の値に正規化している。

式 1 背景的知识に蓄えられた語の総数からみた単語  $i$  の出現率  $x_i$  と、大規模概念ベースから算出した  $idf$  値 ( $idf(i)$ ) [3] を用いて算出した値。「 $x_i \cdot idf(i)$ 」

式 2 語  $i$  の出現率  $x_i$  と候補集合内での出現率  $a_i$ 、記録されている語の種類数  $D$  を用いて算出した値。「 $a_i \cdot \log(\frac{D}{x_i})$ 」

式 3 候補集合内の語  $a$  と現在の話題語  $b$  の関連度 [2]。「 $Rel(a, b)$ 」

式 4 候補集合内の語  $a$  と現在の話題語  $b$  の共起度 [2]。「 $Co(a, b)$ 」

## 2.2.3 選定手法

ここでは優先度を基に話題語群を選定する手法について述べる。選定の方法は様々なパターンが考えられるが、今回提案するのは 4 つである。以下の優先度を使った算出結果から上位 10 個を最終的な話題語群とする。なお、 $R_i$  は式  $i$  における順位で、 $V_i$  は式  $i$  の優先度である。

$$\text{パターン 1 } \frac{1}{R_1 + R_2 + R_3 + R_4}$$

$$\text{パターン 2 } \frac{1}{R_2 + R_3}$$

$$\text{パターン 3 } \frac{V_1 V_2 V_3 V_4}{V_1 V_2 V_3 V_4 + (1 - V_1)(1 - V_2)(1 - V_3)(1 - V_4)}$$

$$\text{パターン 4 } \frac{V_1 V_4}{V_1 V_4 + (1 - V_1)(1 - V_4)}$$

## 3 評価実験

「弁当」、「本」、「犬」に関する背景的知识を予め計算機に付与する。例 1 のように定義した背景的知识に対して、次節以降の検証を行う。

A Development of Automatic Changing Subject System to Provide Suitable Topics for Users Taste.

† Kidai HAYASHI

† Noriyuki OKUMURA

Nagano National College of Technology, Department Electronics and Computer Science, noriyuki.okumura@ei.nagano-nct.ac.jp (†)

例 1: 背景知識の構造

一行文	単語	出現回数
青色の空, 青色の海ですね.	青色	2
	空	1
	海	1

### 3.1 抽出手法の傾向調査

選定手法を考案するために, 式 1~4 の手法の優先度が現在の話題語に対してどのような語を高く評価するのか傾向を調査する.

#### 3.1.1 評価方法

6 種類の単語 (動物, 勉強, 旅行, 新聞, 豚肉, 机) を挙げ, それぞれが現在の話題語としたとき, 式 1~4 での選出結果を被験者に提示する. 現在の話題に対して, 継続した話題が多く含まれる順, 飛躍した話題が多く含まれる順に結果 1~4 を順位付けさせた.

ここで, 継続した話題とは, 現在の話題に似た話題のことであり, 飛躍した話題とは, 現在の話題とは語自体に類似性はないものの, 知識の中で繋がりのある話題のことである.

#### 3.1.2 結果と考察

表 1, 2 は 6 種類の話題語における平均順位, 一位, 二位, 上位になる確率を平均した値をまとめている.

表 1: 観点「継続している話題」の結果 (被験者 21 名)

式	平均順位	一位	二位	上位
1	3.032	17.46%	5.56%	11.90%
2	3.238	0.79%	11.90%	6.75%
3	1.770	45.24%	37.30%	41.67%
4	1.960	36.51%	45.24%	39.68%

表 2: 観点「飛躍している話題」の結果 (被験者 21 名)

式	平均順位	一位	二位	上位
1	1.873	48.41%	32.54%	40.48%
2	1.889	35.71%	46.83%	41.27%
3	3.159	7.14%	8.73%	7.94%
4	3.079	8.73%	11.90%	10.32%

表 1, 表 2 から平均順位が二極化していることが判断できる. つまり結果 1, 2 は飛躍した話題を, 結果 3, 4 は継続した話題を多く含む傾向にあると推察する. そのため, 傾向の異なる語の集合から優先度が上位となる候補を選択することで, より良い話題語群を選出できることが予測される.

次節には以上の結果から選定手法を考慮し, その適切かどうかの調査を行う.

### 3.2 選定手法の調査

選定手法がどの程度適切な話題が候補として含まれているかを調査する.

#### 3.2.1 評価方法

3 種類の話題語 (勉強, 動物, 旅行) を現在の話題語とする. 2.2.3 節に示すパターン 1~4 の手法で得られた話題語群に含まれる全ての語の種類を被験者に提示し, 現在の話題に対し, 相応しくない語を選ばせた.

#### 3.2.2 結果

表 3 は式 1~4 の値を選定手法として使用した場合の結果である. 表 4 は 4 パターンの結果である. それぞれ被験者 9 名によって行われた評価である. 適切な語とは, 相応しくない語ではない, 次の話題として提示すべき語のことである.

表 3: 上位 10 個の内に適切な語が入る確率

式	1	2	3	4
※	71.10%	62.23%	87.43%	87.03%

表 4: 上位 10 個の内に適切な語が入る確率

パターン	1	2	3	4
※	88.89%	87.04%	69.63%	69.63%

※適切な語が含まれる率

## 4 考察

表 4 の結果より, パターン 1, 2 が適切な語が多く含まれる話題語群を選定できていることがわかる. これらの適切な語が含まれる率は表 3 に比べて高く, より適切な語が選定されていると判断できる.

しかし, パターン 1, 2 で選出された話題語群と式 1~4 で選出された話題語群を比較すると, 式 3 で選出された話題語群に類似していた. これは式 1, 2, 4 で話題語群がほぼ選定されていないことを示している. それぞれで選出された話題語群から網羅的に候補を抽出するように, 式 1~4 の複合方法の検討が課題として残る.

## 5 まとめ

本稿では, 背景知識と発話のサイクルにおける「話題決定」の評価を行った. 今後の課題として, 1 つの選出結果に左右されない複合手法の確立が挙げられる.

### 参考文献

- [1] 「自動会話システム (人工無能) の開発とその応用 - Web テキストからの会話文生成と会話形成に関する研究」森部 敦, 毛利 公美, 森井 昌克: IEICE technical report. Office Information Systems 105(283) pp.11-16 20050908
- [2] 「概念の意味属性と共起情報を用いた関連度計算方式」渡部 広一, 奥村 紀之, 河岡 司: 同志社大学理工学研究報告 第 48 巻 第 3 号
- [3] 「概念間の関連度計算のための大規模概念ベースの構築」奥村 紀之, 土屋 誠司, 渡部 広一, 河岡 司: 自然言語処理 Volume14 Number5 p.41-64