

# 応答文生成に向けた内容語選択における精度向上

佐藤 和†  
Kazu Sato

福田 雅志†  
Masashi Fukuda

延澤 志保‡  
Shiho Nobesawa

太原 育夫‡  
Ikuo Tahara

## 1 はじめに

近年、人間と計算機が自由に対話することを目指した研究が行われている。自由な対話とは、対話の内容を制限せずユーザの自由な発話を許容する対話のことである。しかし、自由な対話であっても発話と応答の間に何らかの関連性が無ければ対話は成立しない。即ち、対話を成立させようとする場合には、発話と応答の間に何らかの関連性があることが必要である。そこで佐藤らは、発話と応答の間の関係を連想関係として定義し、発話に対する応答文に用いる語を連想関係に基づいて選択する手法を提案した [1]。佐藤らの手法は連想関係に基づいたスコアリングによって応答文に用いる語を選択する手法であり、インターネット上で行われたチャット対話において、10語を応答文に用いる語として出力した場合、全出力語のうち39.50%の適切な語を出力することができた。また、対話中の過去の発話に含まれる語をスコアリングの際に手掛かりとして用いることで適切な出力語が増加することが示された。

しかし、実験に用いた発話の中には手法の効果が現れない発話も見られた。それらの効果が現れない発話の特徴として、発話が長く発話に含まれる語が多いことが挙げられる。この手法では、過去の発話に含まれる語をスコアリングの手掛かりとして用いる場合、佐藤らの実験において、実験に用いた対話の1発話当たりの平均語数は3.750語であったが、効果が現れなかった発話の平均語数は4.875語であった。また、システムで内容語選択の手掛かりに用いる語数(入力語数)についても全体の平均と問題の発話に対応する入力語の平均を比べると、実験全体での入力語数の平均が12.9250語であるのに対して、問題の発話の入力語数の平均は14.4375語であった。この点から、発話の語数、または入力語数がシステムの精度に影響を与えており、特に、多すぎる入力語はノイズとして働き、出力の精度を悪化させていると考えられる。一定の範囲の発話に含まれる語をすべて用いるため、長い発話が連続した後などは入力語数が非常に多くなってしまふ。そこで、本稿では入力語数を操作することでこの手法の精度を向上する。

## 2 連想に基づく内容語選択システム

対話は対話参加者の間で同一の話題を設定して行われ、その中で交わされる発話と応答の間には話題を介した関連性があると考えられる。発話と応答の間の関連性は、発話に含まれる語と応答に含まれる語の間の関連性として捉えることができる。また、ある物事や概念を表現するために、一定の範囲内で共起する語は、やはりその物事や概念を通じて互いに関連性があり、この関連性は、発話と応答の間の関連性と同類のものであると考えられる。

そこで、ユーザの発話に対する応答文を生成するために、連想辞書と手掛かり語を用いて関連度を基にしたスコアリングを行い、出力候補語群から語を選択するシステムを考える。処理の流れを図1に示す。

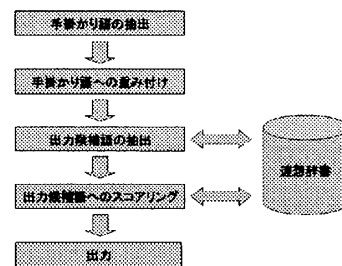


図1: 処理の流れ

### 2.1 連想辞書

連想辞書とは語と語の連想関係を定義した辞書である。連想辞書は見出し語と対応する連想語と重みの対で構成する。図2に連想辞書の一部を抜粋したもの(見出し語“高校”と連想語の一部)を示す。連想語もそれぞれ見出し語として連想辞書に定義されているため、すべての語に対して連想語を連鎖的に定義できる [3]。

ここでは連想辞書をコーパスから作成した。コーパス中の同一の文中で共起する語同士を互いの連想語と定義する。連想語の重みは見出し語ごとに定められており、それぞれの見出し語との共起頻度と共起の偏りを基に定められる。ある見出し語  $w$  の連想語  $w_i$  の重み  $Weight(w_i, w)$  は  $w_i$  が  $w$  とコーパス中で共起する回数  $cof(w_i, w)$  とコーパスに含まれる全ての語の数  $wc$ 、コーパス中で  $w_i$  と共起する語の数  $cov(w_i)$ 、 $w$  と共起

Improvement of Operative Word Selection for Association-based Automatic Reply System.

† 東京理科大学大学院理工学研究科情報科学専攻

‡ 東京理科大学理工学部

1	高校	1.91002
2	学院	0.00223562
3	学期	0.00198595
4	学区	0.00702844
5	学校	0.012884
6	学習	0.00256901
7	学生	0.00157953
8	学部	0.00186272
9	教える	0.00242594
10	教委	0.00426381

図2: 連想辞書の構成

する語の総数  $twc(w)$  を用いて式 (1) のように定義される。

$$Weight(w_i, w) = \frac{cof(w_i, w)}{twc(w)} \cdot \log \frac{wc}{cow(w_i)} + 1 \quad (1)$$

連想辞書はインターネット上の電子掲示板から収集したコーパスを基に作成した。コーパスに対して茶釜 [2] を用いて形態素解析を行い、品詞が名詞、形容詞、動詞の語を見出し語及び連想語として使用した。

作成した連想辞書の見出し語数は 5683 語であり、見出し語 1 つに対する平均連想語数は 32 語である。

## 2.2 関連度

語と語の関連の強さを表現する値として関連度を定義し、連想関係の強さの尺度として出力候補語に対するスコアリングに用いる。ここでは、概念と概念はそれぞれの属性が一致しているほど関連が強いとする渡部らのモデル [3] に基づき語同士の関連度を定義した。語  $w_a, w_b$  の連想語を  $w_{a_i}, w_{b_j}$  とし、その重みを  $u_{a_i}, u_{b_j}$  としたとき、語  $w_a$  と語  $w_b$  の関連度は式 (2) によって定義する。

$$MatchW(w_a, w_b) = \frac{\sum_{w_{a_i}=w_{b_j}} u_{a_i}}{\sum_{i=1}^L u_{a_i}} + \frac{\sum_{w_{b_j}=w_{a_i}} u_{b_j}}{\sum_{j=1}^M u_{b_j}} \quad (2)$$

ここで、 $L, M$  はそれぞれ  $w_a, w_b$  の連想語数である。式 (2) は語  $w_a$  の連想語のうちで語  $w_b$  の連想語と一致しているものの重みの割合と語  $w_b$  の連想語のうちで語  $w_a$  の連想語と一致しているものの重みの割合の平均を表している。

さらに、渡部らは、1次属性を比較する際に属性の完全な一致ではなくその属性が表している概念の一致を利用するという考えに基づいてより高次の関連度を定義している。ここでも同様に、語同士の連想語を比較する際に連想語同士が一致することではなく、連想語同士の  $MatchW$  を利用してより高次の関連度を、式

(3) によって定義する。

$$ChainW(w_a, w_b) = \frac{\sum_{i=1}^L u_{a_i} MatchW(w_{a_i}, w_{b_i})}{\sum_{i=1}^L u_{a_i}} + \frac{\sum_{i=1}^L u_{b_i} MatchW(w_{a_i}, w_{b_i})}{\sum_{j=1}^M u_{b_j}} \quad (3)$$

ただし、 $L \leq M$  であり、 $\sum_{i=1}^L MatchW(w_{a_i}, w_{b_i})$  が最大になるように  $w_{a_i}$  と  $w_{b_i}$  は対応付けられていて、対応にあふれた  $w_{b_j}, j = L+1, \dots, M$  は無視されるものとする。

1次属性だけを用いた  $MatchW$  による関連度の評価では、1次属性が一致しない場合には関連度を与えることができないが、2次属性を用いる  $ChainW$  を用いた関連度の評価では、1次属性が一致しないが、2次属性の一部が一致する場合にも関連度を与えることができ、より詳細な評価が可能である。また、2次属性は1次属性よりも多いため、1次属性だけを用いた場合には一致する属性が無く、関連度の評価が行えないような場合でも、2次属性を用いることで一致する属性が増え、評価が可能となる場合がある。以上の観点より、ここでは関連度として  $ChainW$  を用いた。

## 2.3 手掛かり語の抽出と重み付け方法

本システムでは発話と応答の間に連想関係があると考えて語の選択を行うため、発話に含まれる語を手掛かりにして、出力候補語に対してスコアリングを行う。このとき、応答に用いる語の選択の際に対話の履歴を考慮することで対話の内容に関連性のない語を抑制し、適切な語を選択できることを過去の実験で示した [1]。そのため直前の発話だけでなく、それまでの対話の履歴から抽出した語を用いて出力する語の選択を行う。応答に対する影響の大きさは、それぞれの語に対して重みを与えることで出力語の選択に反映させる。語の重みの与え方は、応答の対象となる発話に含まれる語の重みを 1 とし、それ以前の文脈に含まれる語については、単調減少する関数を用いて重みを与え、複数の発話に出現する語については、各発話で与えられた重みを加算した。そこで、手掛かり語  $w_{h_i}$  の重み  $LoI(w_{h_i})$  を次の式 (4) によって定義した。

$$LoI(w_{h_i}) = \sum_{j=0}^m f(j) \cdot exist(w_{h_i}, S_j) \quad (4)$$

ただし、応答対象となる発話を  $S_0$  とし、以降応答対象となる発話の1発話前の発話を  $S_1$ 、2発話前の発話を  $S_2$  とし、 $j$  発話前の発話を  $S_j$  とする。また、 $f(j)$  は  $j=0$  の時に 1 であり、以降単調減少する関数であり、 $exist(w_{h_i}, S_j)$  は発話  $S_j$  の中に手掛かり語  $w_{h_i}$  が含まれていれば 1 を、含まれていなければ 0 を返す関数である。本システムでは  $f(j)$  として  $\frac{1}{j+1}$  を用いる。

この処理の流れを図3に示す。図3は図の最上段の長方形内ある対話の最後の発話“良くないと思う”に対して応答文に用いる語を生成する際の処理を示している。

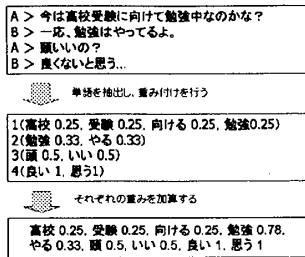


図 3: 手掛かり語の処理の流れ

## 2.4 出力候補語へのスコアリング

出力語は出力候補語群の中からスコアリングにより選択される。手がかり語を  $w_{h_i}$ 、手がかり語  $w_{h_i}$  の重みを  $v_{h_i}$  としたとき、出力語候補語  $w_x$  の重み  $Score(w_x)$  を式 (5) によって定義する。

$$Score(w_x) = \sum_i LoI(w_{h_i}) \cdot ChainW(w_x, w_{h_i}) \quad (5)$$

式 (5) は語  $w_x$  が多くの手掛かり語、また重要な手掛かり語と連想関係にあるほど高い値となる。このスコアリングを行うことで、対話の内容と関連の深い連想語は多くの語との連想関係に由来する高いスコアを与えられることが期待できる。

## 3 余剰手掛かり語の削減

過去の発話に含まれる語を手掛かりに用いることで、出力に含まれる適切な語が増加することが実験により示されている [1]。しかし、過去の発話に含まれる語を手掛かり語に用いた場合でも応答対象の発話に含まれる語だけを手掛かり語に用いた場合と比較して、出力される適切な語は殆ど増加しない例が見られる。このような例は応答対象の発話が長く、多くの語が含まれる場合に多く見られる。この実験で用いた対話の1発話当たりの平均語数は3.750語であったが、実験において3発話前までの発話に含まれる語を用いてスコアリングした場合と応答対象の発話だけを用了場合とを比べて出力語に現れる適切語が増加しなかった、あるいは減少した場合の入力語数の平均は4.875語であった。このことから、原因の1つとして応答対象の発話に多くの語が含まれる場合に、応答対象に含まれる語全てに1の重みが与えられるため、入力語重み全体に対して、応答対象の発話に含まれる語の重みの割合が増大し、過去の発話を用いない場合に近くなることが考えられた。応答対象の発話だけを用了場合と3発話前までの発話に含まれる語を用いてスコアリングした場合とを比べて不適語が減少した数と、応答対象となった発話に含まれる語が手掛かり語重みの総和の中で占める割合との相関関係を調べたところ、その相関係数は-0.415であり、両者には負の相関関係が認められた。この点を考慮し、応答対象の発話に含まれる語の重みの総和を1、 $n$ 発話前の発話に含まれる語の重

みの総和を  $\frac{1}{n+1}$  として実験を行った。その結果を表1に示す。実験の結果は適切な語が1.25%減少した。このような結果になった原因として、応答対象の発話に含まれる語だけでなく、過去の発話に含まれる語の重みも減少したため、全体の語の重みの比率が変化しなかったため、出力語も変化しなかったことが考えられる。この結果は単純に応答対象の発話に含まれる語の重みを制限するだけでは問題が解決しないことを示している。

また、本システムで過去の3発話までに含まれる語を手掛かり語と用いる場合、作成される入力全体の入力語数の平均は12.9250語であった。応答対象の発話だけを用了してスコアリングした場合と、過去の3発話までに含まれる語を手掛かり語としてスコアリングを行った場合とを比較して、発話ごとに適切な語の出力数の増加量を調べ、2名の判定者の平均を取った値が0以下となった場合(16例)の入力語数の平均は14.4375語であった。このことから、一定量を超える入力語は精度を悪化させる要因となると考えられる。そこで、ノイズとして作用する語を入力から削除することによって、適切な出力語の増加が期待できる。以下で入力から語を削除することによる効果を実験により検証する。

表 1: 発話毎の重みの総和を制限した場合の結果

	適切な語 (%)
重み制限なし	39.50
重み制限あり	38.25

## 4 削減された手掛かり語を用いた内容語選択実験

入力語の一部を削除することによって出力語に現れる不適切な語の減少効果を検証する実験を行った。入力インターネット上で行われたチャットを用いて作成し、チャットの各発話に対して応答文に用いる内容語を選択・出力し、入力語の削除を行った場合と行わなかった場合とでその結果を比較し、手法の効果を検証した。

### 4.1 実験の概要

過去の実験 [1] の結果から、本システムにおいても適切な出力語が多かった過去の発話を3発話まで遡って手掛かりとして用いる場合で実験を行った。実験の対象は過去の実験において、過去の発話を用いた場合に過去の発話を用いなかった場合と比較し効果がまったく無かった、あるいは適切な出力語が減少した発話16発話のうち、入力語数が全体の平均より多い12発話である。各発話に対してスコアリング手法によって応答文に用いる語を選択し、スコアの上位10位まで出力した。

## 4.2 実験環境

実験対象となる発話から入力語を1語削除した入力を作成する。入力から1語を削除したすべての場合を検証するため、該当する発話から作成される入力に含まれる入力語数と同数の入力を作成される。その結果、12発話から計189セットの入力を作成した。これらの入力を元に内容語を出力させ、計1890語の出力語を得た。実験結果の評価は出力語が会話の内容や直前の発話に照らして適切かを人手により判断した。

## 4.3 実験結果

入力から余分な語を削除して連想に基づいた内容語選択を行うことによって出力された語を評価した結果を表2に示す。それぞれの出力に対し、出力された10語に対して3名の判定者がそれまでの対話の内容と関連性のある適切な応答文を作ることができるかどうかを判定し、1名の判定者が適切な応答文を作ることができると判定した出力語1語につき、その出力に1点を与え、1つの発話に対してもっとも高得点の出力を、その発話に対する最良の出力であるとした。スコアリング手掛かり語を削除しなかった場合の出力語のうち、3名の判定者のうちの少なくとも2名が適切であると判断した出力語は48語であり、全員が適切であると判定した出力語は6語であった。スコアリング手掛かり語を削除した最良の場合の出力語のうち、3名の判定者のうちの少なくとも2名が適切であると判断した出力語は57語であり、全員が適切であると判定した出力語は14語であった。本実験において、入力から余分な語を削除することで、判定者のうち2名以上が適切であると判定した語は1発話当たり0.750語増加した。そのうち、全員が適切であると判定した語は0.667語増加した。

実験結果より、スコアリングの手掛かり語を削減した最良の場合では、スコアリングの手掛かり語を削減しなかった場合と比較して、3名の判定者のうち2名以上が適切であると判定した語は7.5%増加しており、手掛かり語を削減することで、適切な出力語が増加することが示された。

表 2: 実験結果

	適切な語 (%)
削減しなかった場合	40.00
削減した場合の最良	47.50

## 5 考察

本実験において、削除することで出力語において適切語がもっとも増加する入力語（以下でノイズ語とする）の特徴として、ノイズ語とそのほかのスコアリング手掛かり語の関連度を計算した場合、高い値になることが挙げられる。即ち、1つの入力から1語を取り出し、取り出した語以外の入力語と取り出した入力語

との関連度を計算し、取り出した語以外の入力語すべてとの関連度の平均値が、ノイズ語の場合他の語より高くなっている。入力語同士の関連度の平均値と、ノイズ語と他の語の関連度の平均値を比べた場合、最大で32.0825%、平均で16.5183%ノイズ語と他の語の関連度の平均値が高くなっている。ノイズ語に与えられた重みの平均は0.8556であり、これらの語は大きな重みを与えられていると言える。重みの大きい語であり、そのうえ関連度も大きなスコアリング手掛かり語が存在することで、そのスコアリング手掛かり語が出力候補語に与える重みが大きくなりすぎ、出力語選択に悪影響を及ぼしていたと考えられる。入力語同士の関連度の平均値と、ある語とそれ以外の語の関連度の平均値は事前に知りうる値であり、この2つの値の差を用いることによって、入力語を削減、あるいはノイズ語が与えるスコアを制限することで、精度を向上することが可能であると考えられる。

## 6 おわりに

連想に基づいた応答文生成のための内容語選択において、多くの語を含む入力から余剰な入力語を削減することで出力語に含まれる適切語が増加することを示した。また、余剰な入力語の自動検出・削除に関して入力語同士の関連度の平均値と、各入力語と他の入力語との関連度の平均値の差を利用して余剰な入力語の削除し、精度を向上させる手法の提案を行った。

## 参考文献

- [1] 佐藤和, 延澤志保, 太原育夫, “連想に基づいた応答文生成のための内容語選択,” 情報処理学会第68回全国大会, vol.2, pp.483-484, 2006.
- [2] 松本裕治, 北内啓, 山下達雄, 平野善隆, 松田寛, 高岡一馬, 浅原正幸, “形態素解析システム『茶筌』version 2.3.3 使用説明書,” 2003.
- [3] 渡部広一, 河岡司, “常識的判断のための概念間の関連度評価モデル,” 自然言語処理, Vol.8, No.2, pp.39-54, 2001.