原著論文

対話システムにおけるランダム性を考慮した話題展開手法

三上 佳孝, 萩原 将文

慶應義塾大学大学院

Topic Expansion Method Considering Randomness for Dialogue System

Yoshitaka MIKAMI and Masafumi HAGIWARA

Graduate School of Science and Technology, Keio University, 3-14-1 Hiyoshi, Yokohama-shi, Kanagawa 223-8522, Japan

Abstract : In a non-task oriented dialogue system, it is important that the system can expand various topics for continuing a conversation. At the conventional dialogue systems, topic expansion has not been paid attention. The conventional dialogue systems often use the user's input topic word as a topic word of system's utterance. In this study, we propose a topic word expansion method. Concretely, we extract a topic word from user's utterance and generate embeddings using genetic algorithm and embeddings of the topic word from the user's utterance. As the evaluation function of genetic algorithm, we use neural network that learns topic word correspondence in a dialogue corpus, lexical knowledge and heuristic rules. By using the proposed system, we aim to model ambiguity of topic words and make it possible to expand topic words of conversation integrating plural knowledge. As evaluation experiments, we evaluated the proposed system by subjectivity and verified the effectiveness.

Keywords : Topic word, Non-task oriented dialogue system, Genetic algorithm

1. はじめに

対話システムにおける研究には2つの方向性がある。1つは特定のタスクを解決するという明確な目標を持つタスク指向型の対話システムであり、もう1つは特定の目標を持たず、ユーザーとの雑談などを行う非タスク指向型対話システムである[1]。具体的なタスク指向型の対話システムとしてはApple社の「Siri」やNTT docomo社の「しゃべってコンシェル」などがアプリケーションとして実用化され広く提供されている。一方、非タスク指向型の対話システムはタスク指向型の対話システムと比べ、より広い話題に柔軟に返答出来る必要があるなど、構築において解決すべき課題要素が多く現状として実用化にはまだ遠い状態となっている。

多くの非タスク指向型対話システムでは、ユーザーの入力文から話題語と呼ばれる、発話における最重要単語の抽出処理を行う。例として、ユーザーからの「今日は良い天気だ」という入力を考える。話題語抽出ではこのユーザー発話文中で最も内容的に重要な単語として、この場合は「天気」という単語を取得し対話制御や応答文生成に応用する。これは様々な既存の対話システム[2-6]において行われている処理であり、ユーザーの発話文を解析しシステムの発話文を生成する際に重要な情報として利用される場合が多い。またこの話題語の抽出を前提とした既存研究も多く、稲葉ら[7]は入力を話題語として、その話題語と関連した応答文をTwitterのデータを応用して生成する手法を提案している。

一般的な既存の対話システムでは、この話題語をそのままシステムの話語として利用する場合も多い。より起伏に富んだ小気味良い対話を実現するためにはユーザーが入力し

た話題語に対してシステムが多様な話題を展開することが重要であると考えられる。

また、話題語展開に関しては既存のシステムでも様々な手法が組み込まれており[3-5]、これらのシステムでは一般的に入力の話語に対して変換後の話題語が一意に定まるようなモデルが利用されている。しかし現実世界における実際の対話では話題展開は一意の方向ではなく、ある種のあいまい性を含みながら進行していくのがより自然であると考えられる。このようなあいまい性のモデリングを話題の展開において行った研究は今までに行われていない。本研究ではユーザーの入力文から抽出した話題語を入力として、あいまい性を考慮しつつ、より話が広がる話題語に展開するための手法を提案する。

具体的には、最初にユーザー発話文から話題語を抽出しその話題語の分散表現を獲得する。次に、複数の評価関数を定義した実数値遺伝的アルゴリズムに対して獲得した分散表現を入力し、新たに展開後の話題語の分散表現を生成する。評価関数としては対話データ中の話題語の対応関係をニューラルネットワークによって学習したもの、連想概念辞書[8]中の単語対応関係を学習させたもの、経験則的な処理などを統合して用いる。複数の評価関数を統合して用いることで、データからの傾向と辞書の知識、経験則的な関数など様々な方面の知識を統合して単語を選択することが可能になる。また遺伝的アルゴリズムを用いて単語の選択を行うことで実際の対話中で発生する可能性がある、話題を展開する際に生じるあいまい性をモデリングすることが可能となる。

以下、第2章では関連研究について紹介する。第3章では提案システムの概要と詳細を述べ、第4章では評価実験について詳述する。最後に、第5章でまとめを述べる。

2. 関連研究

関連研究として、中野ら[9]は対話における話題展開に分散表現を用いる手法を提案している。これは単語の分散表現[10]を用いて応答文中の話題語を類義表現に変換することで、ユーザーの入力文における話題語と類似した表現を獲得出来るような手法である。この手法は話題の転換に単語の分散表現を応用しており、話題語の変換自体を目的とはしていない。また吉村ら[11]はシソーラス的構造となっている知識ベースに基づいた連想応答手法の提案を行っている。この手法ではユーザー入力文中の話題語から知識ベースを用いた連想を行うことで応答文を生成している。この手法は知識ベースを用いることで頑健な連想を行っているが、連想の際のあいまい性に関するモデリングには焦点を当てていない。本研究では辞書的な知識も含めた多面的な視点から話題語の転換を行い、あいまい性を考慮した話題の展開を行う手法を提案する。

山内ら[12]は話題を特定の方向に誘導する為の手法を提案しており、それらの挙動をマルコフ決定過程に基づきモデル化している。またIshidaら[13]は現在の話題語と関連した単語をユーザーに提示することで、システムとの対話に慣れていないユーザーの対話補助を行うような機構の提案を行っている。これらの手法ではユーザー補助、あるいはユーザーの興味を持続するための手法として話題語の転換を行っている。

また対話システムに遺伝的アルゴリズムを応用している研究としては木村ら[14]の手法が挙げられる。この手法では対話における応答ルールの獲得に焦点を当てており、話題語や分散表現の変換には焦点を当てていない。本手法では分散表現の変換に複数個の評価関数を用いた実数値遺伝的アルゴリズムを応用する。

3. 提案システム

3.1 提案システムの概要

提案システムの概要を図1に示す。入力文から話題語の抽出を行い、抽出された話題語に対して分散表現のベクトルを獲得する。獲得した入力文の話題語ベクトルを用いて遺伝的アルゴリズム（Genetic Algorithm: GA）による話題語の変換を行い、最後に変換後のベクトルに対して最も類似度の高い単語を出力とする。以下、各モジュールの詳細を述べる。

3.2 話題語の抽出

提案システムではユーザーからの入力文を受け取った後、最初に話題語の抽出を行う。話題語の抽出は既存システム[15]などでも用いられている、各単語のIDF値（Inverse Document Frequency: 逆文書頻度）を比較する手法を用いた。IDF値は各単語 t に対して以下の式で定義される値である。

$$idf(t) = \log\left(\frac{N}{df(t)}\right) + 1 \quad (1)$$

図1 提案システムの概要

ここで、 N は全文書数、 $df(t)$ は各単語 t が出現する文書数を表す。提案システムではユーザーの発話文に対して形態素解析を施し、その結果出力された名詞の中で最もIDF値の高い単語を話題語とした。各単語のIDF値については2016年4月18日時点での日本語版Wikipediaデータを使用し算出した。具体的には、Wikipedia記事中における各章を1文書として扱い、各単語のIDF値を計算した。また形態素解析にはMeCab[16]を用い、辞書としてipadic-neologd[17]を用いた。

3.3 話題語ベクトルの獲得

次に、抽出した話題語を分散表現に変換する。分散表現とは低次元の実数値ベクトルのことであり、言語を数値的に表現する手法の一つである。近年文書分類や機械翻訳など様々な言語処理タスクに応用されており、既存の手法と比べて高精度を記録している[18, 19]。提案システムでは、単語の分散表現をニューラルネットワークによって学習し獲得するプログラムであるword2vec[20]を用いて単語の分散表現を学習した。またword2vecの学習用データにはIDF値の算出と同様に日本語版Wikipediaコーパスを利用した。

3.4 実数値遺伝的アルゴリズム

話題語の分散表現を獲得した後に、評価関数を複数個用いた実数値遺伝的アルゴリズムを用いて変換後の話題語ベクトルを生成する。提案システムでは対話コーパス中の話題語ペアを学習したニューラルネットワーク、連想概念辞書中の単語ペアを学習したニューラルネットワーク、cosine類似度による単語分散表現と出力候補ベクトルとの評価の3つを評価関数として統合して利用した。

一般的に、遺伝的アルゴリズムは離散値の問題に対する最適化を行う際に用いられるメタヒューリスティクスの一つである。しかし、各パラメータの値を実数値に変更することで特定の特徴空間上で探索を行うアルゴリズムとして利用することが出来る[21]。提案システムではこの特徴空間上での探索を分散表現の生成に応用する。遺伝的アルゴリズムは探索結果が初期値に依存しやすい手法であるが、一方である

対話システムにおけるランダム性を考慮した話題展開手法

程度適切な複数個の解が考えられる場合には様々な方面を効率よく探索することが出来るという利点が考えられる。また評価関数が複数個存在する場合には、それぞれの関数をバランス良く満たすように、複数の解候補を同時に比較しながら探索を行うことも可能である[21]。提案システムでは、これら利点をあいまい性のモデリングに応用する。すなわち、実際の対話においても話題語が適切であれば、話題がどのような方向に展開してもある程度許されるという性質のモデリングに遺伝的アルゴリズムを応用する。

提案システムでは、多目的関数の最適化アルゴリズムとしてNSGA II (Non-dominated Sorting Genetic Algorithms- II)[22]を用いた。これは複数個の評価関数に対して最適となる解を探索するためのアルゴリズムの一つであり、既存のアルゴリズムと比べて高速に、かつ大域的に複数のパレート最適解を得ることが出来るという特徴がある。ここで、パレート最適解とは解 x に対して優越する解が存在しない場合の x を指し、「 x_1 が x_2 に対して優越する」とは個体 x_1 と個体 x_2 、 m 個目の評価関数 f_m に関して以下の条件が満たされている場合を指す。

$$\forall m[f_m(x_1) \geq f_m(x_2)] \cap \exists m[f_m(x_1) > f_m(x_2)] \quad (2)$$

図2に具体的なNSGA IIの個体選択アルゴリズムをフローチャートにて示す。図2中の個体の評価では遺伝的アルゴリズムにおける各解候補個体に対してそれぞれの評価関数を用いた評価を行う。また、この評価後に各個体に関して以下の2点を調べ、その値を保存する。

- いくつの個体に優越されているか
- どの個体に対して優越しているか

図2 NSGA IIの個体選択アルゴリズム

ランク r 個体の処理では、その時点で優越されている個体数が0となっている個体群をランク r の個体として1つのまとまりとして個体群から除外する。その後ランク r 個体以外への処理で、ランク r 個体群に優越されている各個体の「支配されている個体数」の値をそれぞれ1ずつ引く。この処理を解候補個体群がなくなり、全ての個体にランクが振られるまで行う。

全個体にランクを振り分けた後、上位ランクの個体から順に優れた個体を選択していく。選択個体数 N を各ランクでの選択個体数の和を超えた場合、選択している最大ランク中から混雑度という指標を用いて選択個体の選定を行う。例として、選択個体数 $N=50$ でランク1の個体数が20、ランク2の個体数が40だった場合を考える。この場合、ランク1の20個体は全て選択され、ランク2の40個体中から混雑度に基づき30個体を選択されることとなる。混雑度はランク中での個体のばらつきを表現する指標であり、隣接している個体との距離を使用する。隣接個体との距離が遠い個体を優先して選択することにより、個体の多様性保持を期待することが出来る。提案システムでは隣接個体との距離算出にcosine類似度を用いた。2つのベクトル $\vec{p}, \vec{q}$ の間のcosine類似度は以下の式で表される。

$$\text{cosine}(\vec{p}, \vec{q}) = \frac{\vec{p} \cdot \vec{q}}{|\vec{p}| |\vec{q}|} \quad (3)$$

具体的な遺伝的アルゴリズムのパラメータ構成としては、提案システムでは以下の値を用いた。300個の実数値ベクトルを初期個体集団として生成し、NSGA IIを用いて優れた100個体を選択、生存させる。また交叉も同様の優れた100個体をコピーして行い、突然変異として世代交代毎に新たに100個のランダムな実数値ベクトルを生成する。提案システムにおいて用いる実数値遺伝的アルゴリズムでの交叉は算術交叉[23]を用いる。2つの個体ペア a, b が与えられた際の算術交叉は以下の式で表される。

$$a' = \lambda a + (1 - \lambda)b \quad (4)$$

$$b' = (1 - \lambda)a + \lambda b \quad (5)$$

ここで、 a', b' はそれぞれ交叉後の新個体を表し、 λ は重みを表す。提案システムでは $\lambda=0.5$ とした。また遺伝的アルゴリズムの終了条件は、NSGA IIのランク付けを行った際に全ての個体がランク1となった際か、世代交代が15回行われた後とした。

3.5 遺伝的アルゴリズムに使用する評価関数

提案システムでは、3種の評価関数を統合して利用する。提案システムにおける3種類の評価関数は提案システムを用いずに話題語変換を行う場合に利用できる手法を応用した。具体的に、提案システムを用いずに話題語の変換を行う場合、一般的な手法としては以下のものが考えられる。

- 特徴量と分類器による機械学習を用いる手法
- 辞書的データなど、知識ベースを用いる手法
- 人手で決定したルールを用いる手法

図3 ニューラルネットワークの構造

上記の3種類の手法を併用するのは一般的には困難であるが、遺伝的アルゴリズムを用いて評価関数という形にすることで3種類の評価関数を同等として扱い、単純な形で併用することができる。提案システムにおいてはそれぞれ、対話コーパスを用いたニューラルネットワークによる機械学習を上記における手法Aと、連想概念辞書による辞書の知識の学習を手法Bと、cosine類似度を手法Cと対応づけている。さらに多くの評価関数を併用することも可能であるが、提案システムのモデルとしての有効性検証という側面を考慮し本研究では一般的に使用される手法と対応した3種類の評価関数のみを用いる。以下、それぞれの評価関数について詳述する。

(1) 対話コーパスを学習したニューラルネットワーク：対話コーパス中に存在する話題語の対応関係を、4層ニューラルネットワークを用いて学習しその出力値を評価関数の値として利用する。具体的なニューラルネットワークの構成を図3に示す。図3の左側が学習時の構造、右側が評価関数として利用する際の構造を表している。対話コーパス中で連続する2発話からそれぞれ抽出した話題語のペアをそれぞれ単語の分散表現に変換し、入力として用いる。学習時のラベルには、実際にその話題語のペアが対話コーパス中に存在する適切なものであれば正例として1を与え、コーパス中に存在していないものであれば負例として0を与える。実際に遺伝的アルゴリズムの中で使用する際には図3の右側に示されているよう

に、遺伝的アルゴリズムの各解候補ベクトルとユーザーの入力文中の話題語を入力としてネットワークの出力値を評価関数の値として利用する。

分散表現を入力として正誤を2値で出力するニューラルネットワークは述語項構造の選択選好性推定[24]やSVOモデルの学習[25]など、自然言語処理のタスクにおいて注目を浴びているモデルの一つであり、分散表現を入力とすることで未学習データに対して柔軟に対応することが期待できる。提案システムにおいては具体的に学習する対話コーパスとして、対話破綻検出コーパス[26]の学習用データを用いた。このコーパスはシステムとユーザーの対話から成り立っており、ユーザーの発話に対してシステムの発話が適当であるかどうかのタグが振られている。提案システムの学習ではタグは利用せず、このコーパスの「システムの発話文に対するユーザーの発話文」を1組のペアとして、話題語の抽出を行い学習の正例用データとして利用した。負例については、正例の入力単語に対して任意の単語をランダムで選択し、その単語を負例として使用した。

ニューラルネットワークの具体的な構成として、中間層数は2層、中間層のノード数は2,000とした。また目的関数は最小二乗誤差を利用し、勾配降下の更新規則にはAdam[27]を利用した。また学習回数については、学習時の誤差と学習回数の関係から決定した。学習回数と誤差の関係を図4に示す。図4から、誤差の現象が完全に終了したと考えられる100epochを学習回数として決定した。

(2) 連想概念辞書[8]を学習したニューラルネットワーク：連想概念辞書とは、人手により構築された大規模な連想関係データの一種である。この辞書中のデータを利用したニューラルネットワークについても、ネットワークの構造や学習方法は図3や(1)で示されているものと同様のものを用いて学習を行った。また連想概念辞書には上位概念や下位概念など、単語間の関連性が複数個定義されているが、提案システムでは各単語に対する類義概念のみを学習データとして用いた。学習回数は(1)と同様に誤差と学習回数から決定した。誤差と学習回数の関係を図5に示す。図5より、学習が完全に終了したと考えられる100epochを学習回数として決定した。

図4 (1)で学習したニューラルネットワークのEpoch数と誤差の関係

図5 (2)で学習したニューラルネットワークのEpoch数と誤差の関係

対話システムにおけるランダム性を考慮した話題展開手法

(3) **cosine類似度による評価**：評価関数として、式(3)で表されるcosine類似度を用いる。これは中野ら[9]の研究などでも示されているように、分散表現が類似した表現を応答文に用いることで上手く話題転換が出来る可能性が示唆されていることから加えた指標である。このような経験則的な関数を評価関数として複数併用することができるため、拡張性が高いのもこの手法の利点であると考えられる。

3.6 出力話題語の選択

遺伝的アルゴリズムを用いた実数値ベクトルの生成後は、そのベクトルと最も類似度の高い分散表現を持つ単語を変換後の話題語として出力する。最終的なNSGA IIによる生成ベクトル100個に対して、それぞれ類似度の高い単語を選択する。この処理を一つの入力単語に対して複数回行い、出力回数上位の単語を変換後の単語として出力する。提案システムでは処理を行う回数を5回とした。この際の類似度計算には式(3)のcosine類似度を用いる。出力する変換後の話題語候補については、あらかじめ作成された出力候補データベース中から出力を行う。提案システムでは対話破綻検出コーパスの学習データに表れている話題語と連想概念辞書中に類義概念として表れている名詞の計4,530語を出力候補の話題語として用いた。

4. 評価実験

4.1 単語の変換に関する主観評価実験

(1) **実験条件**：実際に話題語の変換を行い、その印象を主観的に評価する実験を行った。変換する話題語は対話破綻検出コーパスの評価用データ中に存在する話題語からランダムに30単語を選択した。評価項目としては以下の3項目を利用した。

- ・入力単語と変換語について、共通項が存在するか
- ・入力単語から変換単語が連想可能か
- ・対話における話題の中心語として変換された場合に違和感がないか

これらの評価項目は単語の変換自体の印象を問うことを目的として設定したものである。この3項目に関して変換前の単語と変換後の単語を見比べて{1:全く思わない, 2:思わない, 3:普通, 4:そう思う, 5:とてもそう思う}の5段階で評価を行った。被験者は20代の男性6人、女性4人の計10名である。

(2) **結果と考察**：実験結果を表1に、また実験結果として変換された単語の具体例と評価の平均値を表2に示す。表2には各入力話題語と、それに対する変換の結果が示されている。実際に実験結果として出力されているのはそれぞれの入力語に対して右隣に書かれている単語(「雨」という入力であれば「雪」)であり、その右側に各評価項目の平均値が示されている。またその右側にNSGA IIのアルゴリズムを5回処理して出力された変換後の単語500語のうち、結果として出力する確率が次に高かった上位単語2単語が示されている。

表1中の結果に関しては全ての項目で4.0を上回る結果となった。また入力単語と変換語に関しての共通項が存在するかについては他の2項目よりも高い結果となっており、連想可能かどうかと話題の中心語として変換された場合の適切さは平均の差が0.06と非常に近い値となった。

表2中の結果を見ると、「雨」という入力単語に対して「雪」という出力、「電話」という入力に対して「メール」という出力、「ラーメン」という入力に対して「焼肉」という出力など連想や対話の遷移としても適切と考えられる関係が多いことがわかる。また提案システムは遺伝的アルゴリズムを用いているため入力する度に出力単語が変わる可能性があるが、表2中の出力単語を見る限り確率上位の単語の多くは言い換えとしての外れでないと考えられる。このことから、複数回単語を入力した場合にもある程度適切かつ多様な出力を実現できており、出力単語に関するあいまい性を考慮できていることが示唆される。

入力単語と変換後の単語の共通項については、例えば「水族館」という入力単語に対して「レストラン」は連想として適当であるか、対話中の話題語の遷移として適当であるかは微妙と考えられる。しかし、「場所」という面においては共通している。このようなことから、共通項が存在しているかどうかという質問項目は他の2項目と比べて評価値が高くなったと考えられる。また評価値が満点に至らなかった原因とし

表1 単語の変換に関する主観評価実験結果

評価項目	平均値	標準偏差
入力単語と変換語に共通項が存在するか	4.38	0.56
入力単語から変換単語が連想可能か	4.20	0.62
対話における話題の中心語として変換された場合に違和感がないか	4.14	0.68

表2 単語の変換に関する具体例と評価の平均値

入力話題語	出力話題語	共通項が存在するか	連想可能か	話題語の変換として違和感がないか	出力確率2位の単語	出力確率3位の単語
雨	雪	5.0	4.9	4.5	晴れ	霜
電話	メール	4.5	4.8	4.3	テレビ	通信
ラーメン	焼肉	4.6	3.3	3.9	麵	餃子
水族館	レストラン	3.7	3.0	3.6	動物園	タワー
和菓子	菓子	5.0	4.8	5.0	ラーメン	書籍
ロボット	コンピュータ	4.7	4.6	4.6	ロボ	モンキー
研究	科学	4.7	4.5	4.3	博士	芸術
文化	社会	4.0	4.0	4.0	郷土	経済

では、「ロボット」という入力単語に対して「コンピュータ」という上位概念にあたる、入力単語を包括したような単語の出力があったことが一因であると考えられる。この点に関しては「和菓子」という単語に対しても同様に、「菓子」という上位概念的な単語が出力しているが、連想語や対話における話題の変遷としてさほど違和感がない出力となっている。このようなある種の上位概念語などの微妙な印象の違いが、評価値が満点に至らなかった一因であったと考えられる。

また、「文化」という入力単語に対しては「社会」という抽象的な単語が出力されている。このような抽象的な単語の出力も主観的には評価が難しい点であり、評価値が満点に至らなかったことの一因であると考えられる。

上記のような具体的な事例の考察から、今回の提案システムによる変換において変換名詞が具象名詞か抽象名詞かによって出力の精度が変わると考えられる。具体的な物事を指し示している具象名詞（「雨」、「電話」、「ラーメン」、「水族館」など）に関しては具体的な出力例からも適切な変換が期待できるが、「研究」や「社会」などの抽象的な行為、概念を示す抽象名詞に関しては変換結果が具象名詞と比べて適切になりにくいと考えられる。実際に今回の実験の入力として使用した単語をそれぞれ具象名詞と抽象名詞に分類し、実験結果の平均と標準偏差をとったところ、表3のような結果が得られた。表3の結果より、具象名詞と抽象名詞というカテゴリの違いによって変換結果に差が生じる可能性が示唆される。

また、今回の評価実験において出力された単語は、遺伝的アルゴリズムにおける評価関数のニューラルネットワーク及びコサイン類似度によって最終的な判定が行われている。

表3 名詞を分類した際の単語の変換に関する主観評価実験結果

評価項目	具象名詞		抽象名詞	
	平均	標準偏差	平均	標準偏差
入力単語と変換語に共通項が存在するか	4.57	0.50	4.19	0.62
入力単語から変換単語が連想可能か	4.33	0.54	4.07	0.69
対話における話題の中心語として変換された場合に違和感がないか	4.28	0.67	4.00	0.68

また入力単語の分散表現も word2vec によって生成されており、それぞれ分散表現やニューラルネットワークの性質は学習コーパスに大きく依存するものとなっている。今回分散表現は日本語版 Wikipedia を、ニューラルネットワークは対話破綻検出コーパスと連想概念辞書を学習のコーパスとして使用しているが、これら学習データの性質が出力結果に大きな影響を及ぼしている可能性が考えられる。このようなコーパスの偏りは、特に提案システム中で直接の評価関数として用いられているニューラルネットワークで強く現れやすいと考えられる。具体的に、今回ニューラルネットワークの学習に利用したコーパスの一つである連想概念辞書などでは類義概念のみを学習データとして利用している。このことが表2中の変換結果で類義的な単語の出力が多かったことの一因であると考えられる。表2では上位的な概念の出力も見ることができるが、これは評価関数中のサイン類似度の寄与があったこと、またそのような上位的な単語の分散表現が入力単語の類義語の分散表現と類似していたことなどが理由として考えられる。このような評価関数における学習データを変更することで、今後より多様性の高い変換を行うことのできる可能性が考えられる。

上記のような出力の精度に関するばらつきは考えられるものの、表2中の評価項目の平均では、全ての項目、具体例において数値は3.0以上となっている。また表1の各項目における標準偏差の平均からも、それぞれの評価項目のばらつきは0.5程度となっていることがわかる。このことからどのような単語に対しても評価のばらつきを生むことなくある程度安定した単語の変換が行えていることが示唆される。

また、表4に各被験者の評価平均値と標準偏差を示す。表4において、全ての項目における標準偏差の値が1.0を超えているのは被験者5と被験者10の2名のみであり、それ以外の被験者の標準偏差値は1.0よりも小さい値がほとんどとなっている。またそれぞれの評価項目における被験者毎の平均値に対する標準偏差はそれぞれ0.5よりも小さい値となっている。これらの結果から、どのような被験者に対しても大きなばらつきを生むことなく単語の変換が行えていることが示唆される。これらの実験結果から提案システムを用い

表4 単語の変換に関する各被験者の評価値と標準偏差

	共通項が存在するか		連想可能か		話題語変換として違和感がないか	
	平均	標準偏差	平均	標準偏差	平均	標準偏差
被験者1	4.47	0.72	4.20	0.91	4.20	0.75
被験者2	3.87	0.96	4.17	1.07	4.33	0.83
被験者3	4.87	0.50	4.57	0.76	4.93	0.36
被験者4	4.67	0.70	4.40	0.95	4.63	0.75
被験者5	4.03	1.14	3.87	1.23	3.50	1.06
被験者6	4.20	0.98	4.23	0.88	4.23	0.96
被験者7	4.50	0.81	4.10	1.11	3.80	1.14
被験者8	4.83	0.52	4.57	0.72	4.07	0.89
被験者9	4.47	0.76	4.70	0.64	4.03	0.95
被験者10	3.87	1.15	3.17	1.29	3.63	1.30
平均	4.38	0.35	4.20	0.42	4.14	0.41

対話システムにおけるランダム性を考慮した話題展開手法

ることでどのような被験者、単語に対しても柔軟な単語の変換が行えると考えられ、提案システムによる単語想起、ならびに話題語変換手法としての有効性が示唆される。

4.2 発話文中の話題語変換に関する主観評価実験

(1) **実験条件**: 2種類目の主観評価実験として、実際に対話文中の話題語を変換後のものに置き換え、変換前の印象と比較する実験を行った。こちらもコーパスとして、4.1節と同様の対話破綻検出コーパスの評価用データ中から以下の基準を満たす発話対(ユーザー発話に対するシステム応答文の対) 30対を主観評価用データとして用いた。

- ・ユーザー発話文中の話題語がそれに対するシステムの応答文中にも含まれている
- ・応答文中に話題語が1単語のみしか含まれていない
- ・システム応答文の文長が15文字以内

また、話題語変換前後のシステム応答文を比較する形で評価を行い、以下の4項目を評価項目として用いた。

- ・入力文(ユーザーの発話文)に対する応答文(システムの応答文)の妥当性の高さ
- ・応答文の文としての自然さ
- ・今後話を続ける際の話題展開の可能性
- ・応答文の意外性

これらの評価項目は、話題語を実際に対話システムの応答文として変換した際にユーザーにどのような印象の違いを与えられるか検証することを目的として設定したものである。評価指標についてはそれぞれの項目に対して変換前と変換後と比較して、{1:とても低い, 2:低い, 3:普通, 4:高い, 5:とても高い}の5段階で評価を行った。また、被験者については4.1節と同様の20代男性6名20代女性4名の計10名である。

(2) **結果と考察**: 実験結果を表5に示す。また、対話システムの出力例とその評価平均値を表6に示す。表5の全ての項目の平均値が3.0を超えていることから、既存システムによる発話文と比較した場合には話題語変換を行った方が全体として少なくとも4項目に関しては同等かそれ以上の質の文生成を行えたことが示唆される。具体例として表6より、「私は塩ラーメンが

好きです」という入力文に対して既存の手法では単純な繰り返しを応答していたのに対し、提案システムを応用して単語を置き換えることでより深い話題へ導入するような話題の転換が行われている。また応答文の自然さという点においても表6中の「通勤は電車ですか?」という入力に対して既存システムでは的はずれな応答文を返してしまっているが、提案システムを用いることで意味の通る「電車に乘ります」という発話を行うことができている。「水族館だね」という入力文に対しても同様に、提案システムを用いることで意味の分かる発話を生成することが出来ている。また話題展開の可能性、意外性という面に関しても、従来手法では「野球は好きですか?」という発話に対して「サッカーは良いですね」と返答を返すことで対話の主導権がシステム側に移動し、より展開に富んだ幅の広い対話を行える可能性が示唆される。

応答文としての妥当性に関しては、そもそもの対話破綻検出コーパス中のシステム出力があまり適切といえない文に関して単語変換を行っている場合があるため、他の項目と比較して低い値になっていると考えられる。表6中の「何のアンコール?」という発話文に対する「アンコールを見るのです」という応答文は、どのような置き換えを行っても正しい発話文にならない可能性が考えられる文である。このようなテンプレート選択自体の問題により精度が低下していることは充分に考えられ、これについては別途テンプレート選択手法による解決が必要であると考えられる。

また応答文の意外性に関しては他の項目と比較して高い値となっているが、これは話題語変換による効果であると考え

表5 対話中の話題語変換における主観評価実験結果

評価項目	変換前後の比較結果 (平均値)	標準偏差
入力文に対する 応答文の妥当性	3.41	0.95
応答文の自然さ	3.76	1.04
今後話を続ける際の 話題展開の可能性	3.75	0.89
応答文の意外性	3.88	0.71

表6 対話における話題語の展開例と評価平均値

ユーザーの入力文	出力文 (上段: 変換前, 下段: 変換後)	応答文の 妥当性	自然さ	話題展開の 可能性	意外性
私は塩ラーメンが好きです	塩ラーメンがいですね	3.5	3.6	3.7	3.8
	麵がいですね				
通勤は電車ですか?	通勤に乘ります	4.9	4.9	4.1	3.4
	電車に乘ります				
水族館だね	水族館の方が楽しめますよね	3.7	4.3	4.1	4.4
	博物館の方が楽しめますよね				
野球好きですか?	野球は良いですね	2.3	3.5	3.4	4.3
	サッカーは良いですね				
何のアンコール?	アンコールを見るのです	3.4	4.1	3.5	3.4
	コンサートを見るのです				
梅雨明けましたね	梅雨に突入するのでしょうか?	3.0	3.9	3.5	3.7
	台風に突入するのでしょうか?				

られる。表6の「梅雨明けましたね」という発話に対しては梅雨という単語から台風という単語を変換して出力することにより意味も通りやすく、発話の方向性としても意外性の高い文脈を作りだすことに成功している。

表6の具体例の評価平均値より、提案システムを用いた際の出力はほとんどの場合普通、もしくはより良くなっている場合を示す3.0以上の値となっている。「野球は好きですか?」という入力文に対する変換は応答文の妥当性という項目において表6中で唯一3.0を下回る数値を記録しているが、話題展開の可能性や文としての自然さ、意外性など他の評価項目においては3.0を超える数値を記録している。

また表5中のそれぞれの項目の標準偏差については、応答文の自然さが最も大きい値となっており、それ以外の3項目ではそれぞれ値が1.0よりも小さくまたそれぞれの評価平均値が大きい項目ほど標準偏差の値が小さくなっている。「応答文の自然さ」に関しては、この項目が変換前の文章の質に依存しやすいものであることに由来していると考えられる。提案システムにおいては、新たな話題展開の可能性及び対話における意外性を拡張することを主な目的としており、表5中の標準偏差の小ささや評価平均値の高さからその二項目において提案システムの有効性が示唆される。

また、表7に各被験者の評価値と標準偏差を示す。表7より話題展開の可能性及び意外性の項目に関しては全ての被験者において評価値が3.0を超えており、また被験者6を除いた9名については標準偏差の値が1.0以下となっている。このことから提案システムを用いることで対話中においても、様々な被験者に対して大きなばらつきを生むことなく話題語変換を行える可能性が示唆される。応答文の妥当性に関しては評価者によって平均値のばらつきが大きい傾向があるが、これは前述したテンプレートの選択精度に関する問題以外にも評価項目の解釈に関して個人差があった可能性が考えられる。このような対話システムの評価方法は現在でも確立した手法が存在せず、議論が続けられている分野でもあるため[1]、今後より深い調査が必要であると考えられる。

上記の実験結果や考察より、対話中での話題語変換に関しても提案システムを用いることでばらつきを生むことなく、

意外性の向上やより多様な話題展開を指向した話題語の変換を行うことができると考えられる。

5. お わ り に

本論文では、対話システムにおける話題の展開手法を提案した。具体的には実数値遺伝的アルゴリズムと複数の評価関数を組み合わせて分散表現の変換を行うことで、あいまい性を考慮し未知語や未学習語に対して高い柔軟性を保ちつつより多様な方向へ話題を導けるような話題語の変換を行うシステムを提案した。

主観評価実験の結果より、あいまい性を保ちつつ様々な語に対して頑健な話題語の転換が行えることが示された。また対話文中の話題語を変換して評価実験を行った結果からは実際の対話文中における話題語の変換においても提案システムが有効であることが示唆された。

今後の課題としては、文脈情報や発話行為タグ、構文情報など様々な評価関数を用いた際の性能評価が挙げられる。また今後の展開としては、提案システムを念頭に置いた文生成のためのテンプレート選択手法の提案や連想モデルをはじめとした知識処理分野への応用などが考えられる。

参 考 文 献

- [1] 中野幹夫, 駒谷和範, 船越孝太郎, 中野有紀子, 奥村学: 対話システム(自然言語処理シリーズ), コロナ社, 2015.
- [2] Weizenbaum J.: ELIZA-a computer program for the study of natural language communication between man and machine, Communications of the Association for Computing Machinery, 9(1), pp.36-45, 1966.
- [3] 小林峻也, 萩原将文: ユーザの嗜好や人間関係を考慮する非タスク指向型対話システム, 人工知能学会論文誌, 31(1), pp.DSF-A_1-10, 2016.
- [4] 伊藤直之, 西川侑吾, 大野和久, 松本征二, 中川修: 語の意外度に基づき話題展開する非タスク指向型対話システム, 人工知能学会全国大会論文集, 29, pp.1-4, 2015.

表7 対話中の話題語変換に関する各被験者の評価値と標準偏差

	応答文の妥当性		自然さ		話題展開の可能性		意外性	
	平均	標準偏差	平均	標準偏差	平均	標準偏差	平均	標準偏差
被験者1	4.00	0.73	4.27	0.77	4.20	0.65	4.10	0.70
被験者2	3.53	1.02	4.00	1.00	3.73	0.81	3.83	0.58
被験者3	3.97	0.95	4.17	0.90	4.77	0.50	4.53	0.67
被験者4	4.13	0.99	4.70	0.64	4.77	0.88	4.10	0.94
被験者5	3.17	1.24	3.90	0.94	3.73	0.93	3.80	0.70
被験者6	2.73	1.18	2.87	1.33	3.03	1.30	3.53	1.06
被験者7	2.60	0.84	2.57	0.88	3.33	0.98	3.63	0.71
被験者8	3.23	0.72	3.30	0.82	3.20	0.40	3.77	0.76
被験者9	2.87	0.92	3.00	0.52	3.73	0.68	3.90	0.83
被験者10	3.83	1.00	4.80	0.54	3.27	0.44	3.67	0.75
平均	3.41	0.96	3.76	0.84	3.75	0.76	3.89	0.77

対話システムにおけるランダム性を考慮した話題展開手法

- [5] 大竹裕也, 萩原将文: 評価表現による印象推定と傾聴型対話システムへの応用, 日本知能情報ファジィ学会誌, 26(2), pp.617-626, 2014.
- [6] 柴田雅博, 富浦洋一, 西口友美: 雑談自由対話を実現するためのWWW上の文書からの妥当な候補文選択手法, 人工知能学会論文誌, 24(6), pp.507-519, 2009.
- [7] 稲葉通将, 神園彩香, 高橋健一: Twitterを用いた非タスク指向型対話システムのための発話候補文獲得, 人工知能学会論文誌, 29(1), pp.21-31, 2014.
- [8] 岡本潤, 石崎俊: 概念間距離の定式化と既存電子化辞書との比較, 自然言語処理, 8(4), pp.37-54, 2001.
- [9] 中野哲寛, 荒木雅弘: 雑談対話システムにおける単語分散表現を用いた話題展開手法, 言語処理学会第21回年次大会発表論文集, 1, pp.269-272, 2015.
- [10] Levy O, Goldberg Y.: Dependency-Based Word Embeddings, Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 2, pp.302-308, 2014.
- [11] 吉村枝里子, 芋野美沙子, 土屋誠司, 渡部広一: 知的会話処理における連想応答手法, 人工知能学会論文誌, 28(2), pp.100-111, 2013.
- [12] 山内祐輝, 戸田智基, 中村哲: 対話システムにおける用語間の関係性を用いた話題誘導応答文生成, 人工知能学会論文誌, 29(1), pp.80-89, 2014.
- [13] Ishida Y., Uchiya T., Yamamoto K., Yamamoto D., Nishimura R., Takumi I.: Related Word Recommendation Mechanism for Speech Dialogue System, Proceedings of the 19th International Conference on Network-Based Information Systems, pp.464-468, 2016.
- [14] 木村泰知, 荒木健治, 桃内佳雄, 柄内香次: 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処理手法, 電子情報通信学会論文誌, J84-D2(9), pp.2079-2091, 2001.
- [15] Sugiyama H., Meguro T., Higashinaka R., and Minami Y.: Open-domain Utterance Generation for Conversational Dialogue Systems using Web-scale Dependency Structures, in Proceedings of the Special Interest Group on Discourse and Dialogue 2013 Conference, Metz, France, pp.334-338, 2013.
- [16] Kudo T., Yamamoto K., Matsumoto Y.: Applying conditional random fields to Japanese morphological analysis, in Proceeding of the 2004 Conference on Empirical Methods on Natural Language Processing, pp.230-237, 2004.
- [17] Sato T.: Neologism dictionary based on the language resources on the web for mecab, <https://github.com/neologd/mecab-ipadic-neologd> (2017/10/14 閲覧).
- [18] Le Q., Mikolov, T.: Distributed representations of sentences and documents, Proceedings of the 31st International Conference on Machine Learning, pp.1188-1196, 2014.
- [19] Cho K., Van Merriënboer B., Gulcehre C., Bahdanau D., Bougares F., Schwenk H., Bengio Y.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078, 2014.
- [20] Google Code Archive word2vec: <https://code.google.com/archive/p/word2vec/> (2017/09/27 閲覧).
- [21] 棟朝雅晴: 遺伝的アルゴリズム - その理論と先端的手法 -, 森北出版, 2008.
- [22] Deb K., Pratap A., Agarwal S., Meyarivan T.: A fast and elitist multi-objective genetic algorithm: NSGA-II, IEEE Transaction on Evolutionary Computation, 6(2), pp.182-197, 2002.
- [23] Michalewicz Z.: Genetic Algorithms + Data Structures = Evolutionary Programs, Springer Science & Business Media, 2013.
- [24] 大野雅之, 井之上直也, 松林優一郎, 岡崎直観, 乾健太郎: 分散表現による文脈情報を用いた選択選好モデル, 言語処理学会第22回年次大会発表論文集, 1, pp.885-888, 2016.
- [25] Van de Cruys T.: A Neural Network Approach to Selectional Preference Acquisition, in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, pp.26-35, 2014.
- [26] Higashinaka R., Funakoshi K., Kobayashi Y., Inaba M.: The dialogue breakdown detection challenge: Task description, datasets, and evaluation metrics, In Proceedings of the Tenth International Conference on Language Resources and Evaluation, pp.3146-3150, 2016.
- [27] Kingma D., Ba J.: Adam: A Method for Stochastic Optimization, arXiv preprint arXiv:1412.6980, 2014.

三上 佳孝 (非会員)

2016年 慶應義塾大学理工学部卒業。2018年 同大学大学院開放環境科学専攻修士課程修了。在学中非タスク指向型対話システムの研究及び自然言語処理の研究に従事。

萩原 将文 (正会員)

1982年 慶大・工・電気卒。1987年 同大学院博士課程修了。工博。同年同大助手。現在、同大教授。1991-92年度スタンフォード大学訪問研究員。視覚・言語・感性情報処理とその融合の研究に従事。1990年 IEEE Consumer Electronics Society論文賞, 1996年 日本ファジィ学会著述賞, 2004年, 2014年 日本感性工学会論文賞, 2013年 日本神経回路学会最優秀研究賞受賞。2015-2016年 日本知能情報ファジィ学会会長。