HOKKAIDO UNIVERSITY

Title	汎用的な対話システムにおけるドメイン検出のための複合的手法
Author(s)	谷津, 元樹; ジェプカ, ラファウ; 荒木, 健治
Citation	ファジィシステムシンポジウム講演論文集, 28, 420-425
Issue Date	2012-09-12
Doc URL	http://hdl.handle.net/2115/63613
Type	proceedings
Note	第28回ファジィシステムシンポジウム(28th Fuzzy System Symposium). 2012年9月12日～14日．名古屋工業大学, 名古屋市.
File Information	FSS28-2012_420-425.pdf

Instructions for use

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

汎用的な対話システムにおける ドメイン検出のための複合的手法

Combined Domain Detection Methods in General-purpose Dialogue System

○ [1] 谷津　元樹，　[2] ジェプカ　ラファウ，[3] 荒木　健治
○ [1]Motoki Yatsu, [2]Rafal Rzepka, [3]Kenji Araki
[123] 北海道大学大学院 情報科学研究科
[123] Graduate School of Information Science and Technology, Hokkaido University

Abstract: In this paper, we propose implicit and explicit utterance generation models and a dialogue system in which such methods are implemented. Modularizing classifiers enables the agent to give input utterance tags of multiple features including types of sentences and mood expressions. In the implicit way, the features gotten from the inputted sentences define an agent's internal state. A relativity vector to each domain is sustainably computed based on similarity in Japanese WordNet ontology as the system's internal state. The explicit answers are generated if the input text is classified in Question-Answering domain based on the tags given by classifier modules. In other cases of the classification, the system generates open-domain utterances. We will discuss the result of experiments intended to show characteristics of both domain detection methods.

1　はじめに

ヒトの言語能力を再現する手段として，計算機上に対話システムを構築することが考えられる．対話システムは，使用者の発した発話文の入力に対して何らかの処理を加え，音声又は画面表示などの出力を行なうシステムである．対話システムとの対話は，ある役割 (タスク) の達成を目的とした対話と，そうでない対話とに大別できる．前者をタスク指向型対話，後者を非タスク指向型対話と呼称する．

1.1　タスク指向型対話

タスク指向型対話を行なうシステムの一例として，音声対話処理技術を利用した観光案内システム [1] が挙げられる．タスク指向型対話システムは，入力発話に対して当該システムのもつタスクが属するドメイン (対話のトピックの範囲) にのみ属する応答を出力する能力をもつ．

1.2　非タスク指向型対話

一方，非タスク指向型対話は話者がタスクの達成を考慮しない対話である．非タスク指向型対話を行なうシステムは最近その研究の裾野を広げており，その端緒となる研究は Weizenbaum による ELIZA [2] である．

ELIZA は，機械により応答発話生成ルールを用いて人間 (エージェント) としての応答発話を生成することを可能にした最初の例である．ELIZA 型のシステムでは，人間の発するあらゆる種類の発話に対し何らかの返答を準備した応答発話生成ルールセットが設計可能である．

非タスク指向型対話は，発話が属するドメインが変遷するため，一般にオープンドメイン対話，すなわち対象とする分野を指定しない対話に包含される．

1.3　汎用的な対話システムの必要性

タスク指向型対話システムの対話は，利用を開始する前，タスクを達成する後の状況との接続性や連続性を考慮することができない．また，使用者の要求するシステムの動作 (タスク) がシステムへの入力において明示的であることが実用の前提となる．使用者による発話の大部分が明示的な要求の表出 (命令, 依頼) を占めかつタスク (近隣の名所の紹介，経路案内) に結びつけられたドメイン以外に属する対話や，話題ドメインの遷移頻度の高い対話において，この前提が障害となることが予想される．

1.4　対話の 2 状態モデル

上記の問題を緩和するため，人間の対話との類似に基づき対話モデルを設定した [3]．対話モデルにおいては，汎用的なシステム内で，使用者の目的の明確化・曖昧化に伴いオープンドメインと特定ドメインとの間の状態遷移が行われる．人間による対話の例を表 **1** に挙げる．この対話では，話者 A,B が話題の自由な対話を行なっており，B が A の行いたいことを推理し，対話の末尾において提案している．

表 1: 人間による対話の例

A_1	寒い
B_2	寒いな
A_3	午後には晴れて暖かくなるそうだけど
B_4	そうなんだ
A_5	とにかくなにか暖まるものが欲しいね
B_6	コーヒーでも飲む？
A_7	ああ，それいいね
B_8	じゃあ，喫茶店に行くか

自然な対話の例として挙げられる表 1 の対話において，B_6 のような応答発話は，エージェントの内部の状態または知識に依存したドメインの判定・選択を行なう話者主導型のものである．2 状態の間の遷移の条件として，このようなオープンドメイン対話においてドメインの選択を促進する発話が必要であるといえる．これまでの対話エージェントの研究において，システム主導型のドメイン選択発話を行う手法を著者は確認していない．

1.5 本稿の目的と構成

本稿では，表 1 における発話 B_6 と同種の発話をドメイン選択発話と呼称する．本研究の目的は，オープンドメイン対話から特定ドメインへの移行を，対話の連続性を保持しつつシステム主導により行うことである．本稿では，次の 2 つの手法について提案する．本稿ではははじめに，非タスク指向型対話システムおよび質問応答システムが含まれるモジュール対話システムの詳細を **2** で述べる．**6** では，汎用的対話生成手法の定性的評価のために行なった実験 (発話分類精度評価，印象評価実験) について述べ，**7** で結果を考察する．

また，提案した汎用的対話機能において使用者の関心のあるドメインの選択を行なうためのフィルタ，およびドメイン選択発話生成モジュールについて **5** で述べる．ドメイン選択精度の評価を **6** で行い，その精度に対する考察を **8** で述べる．

2 汎用的対話生成手法

1.3 に述べた考え方を具現化する手法について述べる．汎用的対話機能は，オープンドメイン対話の中で使用者が関心を持っていることを明示化した話題に対しての応答だけでなく，使用者とシステムとの対話履歴から暗黙的に対話の所属するドメインを決定し，使用者の意図に最も近い発話の生成能力をもつことを目的とする．次に列挙する手法が考案されている．

2.1 発話集合からの応答選択ルールによる発話生成

1.2 で述べた ELIZA などが用いる手法では，検索パターンと応答発話テンプレートからなる有限個の応答生成ルールセットを用い，応答発話稿補集合から直接応答発話を選択する．この手法のもつ課題として，応答生成ルールの大きさが発話生成にとり最終的な制約となることがあげられる．

しかし，単一の入力発話パタン集合からの検索に基づいて応答を生成することから他の手法に比べ高速であり，文脈を考慮した統計的発話選択手法 [4] の非タスク指向型システムへの応用もなされている．

2.2 入力発話の特徴に基づく応答発話生成手法の選択

発話生成手法をモジュール化し，それらの手法から入力発話に対し適切なモジュールを選択するためのルールないし選択条件を準備する．この手法を用いて，明示的なドメインに特化した発話生成手法を選択することが可能である．本手法ではドメインに適応した複数の応答生成ルールセットが用いられうるため，出力発話のオープンドメイン発話における話題に対する被覆率の向上が考えられる．

3 応答発話選択における外部モジュール

上記 **2** に述べた機能を有する対話システムを実装した．ドメインに対する明示的な情報を発話文中に含む発話の理解を，以下のモジュールにより構成される機能により行なう．提案システムは，入力発話理解部および発話生成部を複数の外部モジュールから構成される．発話生成，入力発話に対するタグ付与 (フィルタ)，および出力発話に対するスコア付与のための各モジュールを使用する．以下に各モジュールの役割を示す．

3.1 フィルタモジュール

タグに対しスコアを付与する．具体的には，入力発話に対し最適なタグと事前に人手で決定された一定のスコアを出力し，発話の特徴に対するスコア付与に用いる．スコアは，異なるフィルタ間の優先順位の決定に用いることができる．タグとスコアの関係については **4** を参照されたい．

3.2 応答選択モジュール

今井ら [5] による 3 手法 (人対人チャットログとの類似度による評価，および入力文と各候補中の単語の共起に基づく関連性評価) により，同一タグの付与された発話モジュールによる応答発話にスコアを付与する．

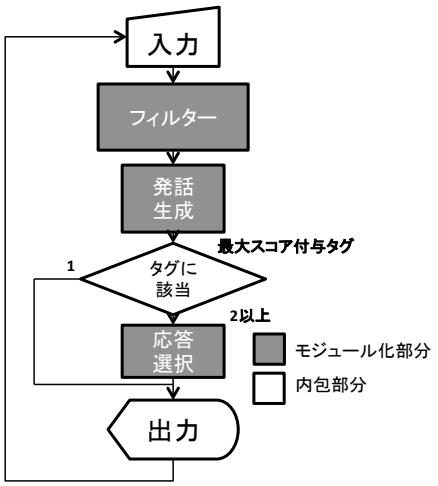

図 1: システムの概要

3.3 発話生成モジュール

モジュール化された対話システムであり，1 入力発話について 1 応答発話を出力する．他の入力先を伴うモジュールもある．対話の種類 (タスク指向型／非タスク指向型，システム主導型／使用者主導型) に関わらず，システムによる応答発話を出力する．各発話モジュールには 1 つ以上のタグと呼ばれる ASCII 文字列 (**4** にて詳述する) が設定されている．

3.3.1 質問応答発話生成 (タグ QA)

Web 検索を知識源とする質問応答システム [8] を用いた．

3.3.2 非タスク指向型応答発話生成 (タグ NT)

本研究では，発話選択モジュールの機能を試験するため，複数の発話生成手法 [6][7][9] を用いた．

3.3.3 ドメイン選択応答発話生成 (タグ RU)

使用者発話に含まれる自立語表現から適切なドメインを選択し，そのドメインキーワードを提示する応答発話を生成するモジュールである．詳細を **5** にて述べる．

3.3.4 タスク指向型応答発話生成 (タグ Dn)

5.1 節で述べるキーワード集合 a_n に対応するタスク指向型応答生成手法である．**6** 節の評価実験では，モデル実装として 1 ドメインに対して 1 タイプの応答発話生成が可能な発話生成モジュールを用いた．

4 外部モジュールの呼出とタグ

システムは入力発話に対する応答を次のような手法で出力する．メインプログラムは，図 1 に示すアルゴリズムに従って，使用者の入力発話から最終的な出力発話を生成する．初めに，入力発話をフィルタモジュールが評価する．戻り値のタグとスコアは，連想配列において加算される．最大スコアの付与されたタグを決定する．

システムは，最大スコアの付与されたタグに属している全ての発話モジュールからの応答を取得する．ここで取得する応答発話が複数となる場合，全ての応答に対し一つまたは複数の応答選択モジュールが評価を行なう．スコアの線形和を比較し，最大スコアを持つ応答発話を出力する．応答発話が単一である場合は，その応答を採用する．

5 ドメイン選択発話生成手法

本稿で提案するシステムにおいて，**1.4** の対話の 2 状態モデルに基づくドメイン選択発話生成手法を提案する．提案手法はオープンドメイン対話における使用者発話と各選択対象ドメインの持つキーワード集合との関連度，使用者の関心の明確度を下記の手法により計算する．閾値と最大値の条件からドメインを選択し，応答発話の生成により使用者に提示する．

5.1 選択対象ドメインとキーワード集合

提案手法の選択対象ドメインは，**2** で提案した汎用的対話機能において実装されたタスク指向型発話生成手法に関連付けられた所与のドメインである．ドメインは，キーワード集合 K_i を持つ．要素であるキーワードは，本システムに組み込まれたタスク指向型発話モジュールと設定ファイルにおける設定により関連付けられた日本語の一般名詞であり，選択対象のドメインに属する．キーワードは，語彙概念辞書の日本語 WordNet [11] における概念でなければならない．

5.2 目的ベクトル $\boldsymbol{a}$

目的ベクトル $\boldsymbol{a} = (a_1, a_2.a_3, ...)$ は使用者の現在の目的とその方向性 (aim) を表す．ここで，ベクトルの要素 a_i は選択対象ドメインのキーワード k_i と発話文中の内容語との関連性を示す．$\boldsymbol{a}$ のユークリッドノルム $\|\boldsymbol{a}\| = \sqrt{\sum_k a_i^2}$ を求めることにより，使用者の目的の明確度を得ることができる．$\|\boldsymbol{a}\| > T$ となるターンにおいて，キーワード集合内のあるタスク指向型発話生成手法に関連付けられたドメインへの関連度が十分に高いという条件により，ドメイン選択発話を生成する．**6** の評価実験においては，$T = 3.0$ とした．

5.2.1 WordNet における意味類似度

$\boldsymbol{a}$ は 意味類似度ベクトル $\boldsymbol{\Delta a}$ の累加によって算出される．$\boldsymbol{\Delta a}$ の要素は Leacock-Chodorow [10] による

意味類似性の尺度によって求められ，次式で表される．

$$\Delta a_i = \frac{1}{N_U N_{K_i}} \sum_{u \in U} \sum_{k \in K_i} sim(u,k) \quad (1)$$

$$where \quad sim(c_1, c_2) = max\left(-log\frac{N_p}{2D}\right) \quad (2)$$

ここで，N_p は 2 単語 c_1, c_2 間の語彙概念辞書における最短のグラフ距離を, D は語彙概念辞書における両者の分類の深さを表す．意味類似度の算出には，Python[1] 上で動作する Natural Language Toolkit[2] を用い，語彙概念辞書に日本語 WordNet を用いた。

5.2.2 使用者の目的の明確度および選択発話生成

目的ベクトル

このとき，フィルターモジュールはタグ 'RU' と正のスコアを出力し，選択される発話生成手法の候補がただ 1 つとなる．発話モジュールは，$\boldsymbol{a}$ における最大の値を持つ要素に関連付けられたキーワード k_i を用い，使用者にドメインの選択を促す発話を生成する (表 7 の発話例参照).

6 対話評価実験

6.1 対話印象評価実験

2 にて提案した汎用的対話生成手法が被験者に与える印象および **5** でのドメイン選択発話生成手法の被験者に対する適性を調査するために行った Web を用いた実験と，アンケート調査による評価について述べる．本研究のベースラインとして，質問応答システム [8] 単体，ELIZA 型システム [9] 単体との比較を併せて行なった．

6.2 実験方法

被験者とシステムとの対話は Web サイト[3]上の提案システムに接続されたスクリプトを用いて，予備実験からドメイン選択発話の生成の予期される 20 ターン以上の自由な対話を行なった．20 ターン経過後に提示されるアンケート画面にて被験者は下記の調査項目について回答した．記録した被験者の属性と分布は、男性 9 名，女性 1 名，職業は研究者・学生以外 4 名，人工知能・自然言語処理に関連する分野の研究者または学生 3 名，人工知能・自然言語処理に関連しない分野の研究者または学生 2 名，小・中・高生 1 名であり，年齢層は 19 歳以下 1 名，20-29 歳 8 名，60 歳以上が 1 名である．

[1]http://www.python.org
[2]http://www.nltk.org
[3]http://ark4.media.eng.hokudai.ac.jp/~my

6.3 印象評価項目

被験者は表 2 に示した項目を用いて印象評価を行なった．ベースラインとして ELIZA 型対話システムにて行われた同様の実験の評価値を参考としてあわせ，結果を表 3 に示す．

表 3: 印象評価実験の結果 (5 段階評価)

項目	評価値平均
A	1.75
B	1.75
C	1.63
D	2.00
E	2.13
F	1.75
提案手法平均	1.85
ベースライン平均	2.54

6.4 形容詞対項目

被験者は，Osgood らによる意味差判別法 [12] での評価に頻繁に用いられているとされる 35 種類 (一部重複がある) の印象を表す形容詞対 [13] を用いシステムとの対話に対し 5 段階の評定を行った．

6.5 発話選択精度評価実験

提案手法における汎用的対話能力の有効性を計量的に示すため，発話生成手法の選択精度の評価を行った．被験者はシステムと使用者の対話について，入力発話も対する応答発話の意味の妥当性を評価する．具体的には，応答発話内容の主観による内容理解可能性，文脈的容認度を 0~3 の評価値として採用する．

6.6 実験方法

被験者は，4 名の 24,25 才の男性理系大学院生である．評価対象とする対話は，**6** に述べた実験にて収集された 150 ターンの対話である．評価は Web を利用して行い，評価対象対話と評価フォームを表示し集計する CGI プログラムを用いて評価データを収集した．評価値及び尺度は，表 4 の通りである．

6.7 結果

選択精度評価の結果を表 5 に示す．全評価の平均値は 2.274，標準偏差は 0.847 であった．

6.8 対話例

印象評価実験で得られた対話の例を，表 7 に示す．

表 2: 印象評価項目

項目	説明
(A)	対話を続けたいかどうか
(B)	対話が文法的に自然であるか
(C)	対話が意味的に自然であるか
(D)	システムの語彙が豊富かどうか
(E)	システムが人間らしいか
(F)	システムによる推薦が，被験者の考えていたことや悩みに対して関連が強かったか

表 4: 選択精度評価実験の 3 段階評価値と尺度

ラベル	値	文法的容認度	意味的容認度
正応答	3	高	高
準応答	2	高	低
誤応答	1	低	低
エラー	0	0	0

表 5: 応答精度評価値 (3 段階評価) の分布

正応答	準応答	誤応答	エラー
302	176	104	17
50.4%	29.4%	17.4%	2.84%

7 印象評価及び発話選択精度に対する考察

実験では，汎用的対話機能およびドメイン選択発話生成モジュールにより生成された発話を評価した．入力がフィルタ機構による分類が正しく行われたものに正応答評価がなされる傾向があったことが，表 6 に示した発話タグ毎の発話精度評価平均値から読み取れる．質問応答発話の精度においては標準的なものが得られ，また非タスク指向型対話システムの中で標準的と捉えられている ELIZA 型対話システムとほぼ同精度で質問応答機能を対話システムに付加することが可能である点において，本提案手法における汎用的対話機能が有効であると考えられる．

7.1 SD 法評価値に対する考察

この項目から得られた評価値データの因子分析を R[4]を用いて行なった．因子分析を行なった結果，実験に使用した形容詞対の群からは 4 つの因子を抽出した．

[4]http://www.r-project.org/

表 6: 発話タグ毎の発話精度評価平均値.

タグ	評価平均値
NT	2.32
QA	2.27
提案手法による全発話	2.28
質問応答システム単体*	1.85
ELIZA 型システム単体*	2.39

これらの因子の中で、「責任感のある-無責任な」「速い-遅い」を含む一群が機能的因子を表していると考えられる．したがってこれらの因子は，印象評価実験においてベースラインシステムと比較して評価が低くなったことの原因としてシステムの処理全体に時間を要していることを示している．ログにおける記録された応答生成の平均所要時間は，ベースライン 0.163 秒に対し，提案手法システム 33.10 秒である．

8 ドメイン選択発話生成の適切性評価項目に対する考察

印象評価における評価項目 (F) はドメイン選択発話生成の適切性を問う項目である．(F) では，ドメイン選択発話がシステムにより行われなかった被験者による平均評価値 (1.83) が低くなる (2.0) 傾向がみられた．

9 おわりに

本稿では，機械と人間との間の自然な HRI の実現に向け，汎用的対話機能においてドメイン選択発話の生成を行う手法を開発した．対話システム全体が被験者に与える印象評価実験では，システムの応答速度が今後の主要な改善点となることが示された．発話選択精度評価実験においては，質問応答システム単体との比較を行なった場合の，本システムにおける汎用的な対

表 7: 対話例．User は被験者，QAC4[8],Recom,Maru[6],Eliza[9] は発話生成モジュール名を表す．

話者	ターン数	発話内容
User	1	日本の首相は誰ですか
QAC4	1	野田佳彦だった
User	2	アメリカの大統領は誰ですか
QAC4	2	オバマらしいですからねぇ
User	3	今日は食堂で夕食を食べる予定です。
Recom	3	もしかして、健康, 病気に興味ありますか？

話機能による発話選択の有効性が示された．

今後の課題として，応答速度および明示的なドメイン選択精度の向上，また暗黙的なドメインの決定における検知可能ドメインの一般化および検出精度の向上が挙げられる．

参考文献

[1] Teruhisa Misu, Komei Sugiura, Tatsuya Kawahara, Kiyonori Ohtake, Chiori Hori, Hideki Kashioka, Hisashi Kawai, and Satoshi Nakamura. Modeling spoken decision support dialogue and optimization of its dialogue strategy. *ACM Trans. Speech Lang. Process.*, Vol. 7, pp. 10:1–10:18, June 2011.

[2] Joseph Weizenbaum. Eliza—a computer program for the study of natural language communication between man and machine. *Commun. ACM*, Vol. 9, pp. 36–45, January 1966.

[3] Motoki Yatsu, Rafal Rzepka, and Kenji Araki. A modules-based, task-navigational dialogue system. *Proceedings of PACLING 2011*, July 2011.

[4] 平井 尚樹, 稲葉 通将, 鳥海 不二夫, 石井 健一郎. 対話エージェントのための文脈を考慮した統計的発話選択. *Proceeding of JAWS2010*, October 2010.

[5] 今井 健太, ジェプカ ラファウ, 荒木 健治. 複数の対話システムからの応答候補文を用いた最適応答文選択手法の性能評価. *IPSJ SIG Technical Report*, Vol. 2010-NL-195(10), pp. 1–7, October 2010.

[6] 高橋 瑞希, ジェプカ ラファウ, 荒木 健治. Web 検索と単語 n-gram モデルを用いた文生成手法の性能評価. 言語処理学会第 16 回年次大会発表論文集, pp.391-394, March 2010.

[7] Shinsuke Higuchi, Rafal Rzepka, and Kenji Araki. A casual conversation system using modality and word associations retrieved from the web. *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, October 2008.

[8] 亀山 恵祐, 荒木 健治, 木村 泰知. 質問応答システムにおける知識源選択規則の自動獲得の有効性について. 情報処理学会研究報告. 自然言語処理研究会報告 35, pp. 85–90, 2007.

[9] 木村 泰知, 荒木 健治, 桃内 佳雄, 栃内 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, Vol. 84, pp. 2079–2091, 2001.

[10] Claudia Leacock and Martin Chodorow. Combining Local Context and WordNet Similarity for Word Sense Identification. *An Electronic Lexical Database*, pp. 265–283, 1998.

[11] F. Bond, H. Isahara, S. Fujita, K. Uchimoto, T. Kuribayashi, and K. Kanzaki. Enhancing the Japanese WordNet. In *Proceedings of the 7th Workshop on Asian Language Resources*, pp. 1–8. Association for Computational Linguistics, 2009.

[12] Charles E. Osgood, George Suci, and Percy Tannenbaum. *The Measurement of Meaning*. University of Illinois Press, June 1957.

[13] 井上 正明, 小林 利宣. 日本における SD 法による研究分野とその形容詞対尺度構成の概観. 教育心理学研究, Vol. 33, No. 3, pp. 253–260, 1985-09-30.