5Q-6

マルチドメイン音声対話システムにおける システム想定外発話のトピック推定に基づく発話の誘導

池田 智志 † 駒谷 和範 ‡ 尾形 哲也 ‡ 奥乃 博 ‡

† 京都大学 工学部情報学科 ‡ 京都大学大学院 情報学研究科 知能情報学専攻

1. はじめに

近年，バス運行案内情報やレストラン・ホテル検索，観光案内などのシングルドメインの音声対話システムをサブシステムとして統合したマルチドメイン音声対話システムが作られている [1]．マルチドメイン音声対話システムでは，システムの扱う発話が多岐にわたるため，ユーザの発話が多様となる．それゆえ，ユーザの発話がシステムの受理できない『システム想定外発話』となることも多い．また，ユーザの発話がシステム想定外発話であると検出できた場合でも，単に棄却するだけでなく，どのサブシステムに制御を戻すのが適当かを決定する必要がある．

本研究ではマルチドメイン音声対話システムにおいて，システム想定外発話を含む，ユーザの多様な発話に対する適切な対話管理の実現を目的としている．すなわち，システムが受理・解釈できない発話が入力された場合でも，Latent Semantic Mapping (LSM) [2] を用いてその発話に最も近いドメインを推定する．これにより，対話の制御をそのサブシステムで行いながら，ヘルプを提示して当該ドメインへと誘導するなどの対処が可能になると期待される．

2. 発話誘導のためのトピック

本研究では，マルチドメインシステム中の１つのサブシステムが受理・解釈できる発話の範囲を **“ドメイン”** と定義する．つまり，あるサブシステムの言語理解部が文法ルールで記述されている場合，その文法で受理可能な範囲がドメインとなる．マルチドメインシステムのいずれかのドメインに含まれる発話は,『システム想定内発話』とする．

一方，ユーザの発話は多様であり，システム想定外発話がしばしば入力される．システム想定外であっても，あるドメインの内容を意図した発話の集合を **“トピック”** と定義する．トピックを定義することで，ユーザがあるドメインを意図したにもかかわらずシステム想定外となった発話に対しても，単に棄却するのではなく，最も近いドメインとしてユーザの意図を推定できる．システム想定外発話に対してトピックを推定することで，ユーザ発話をシステム想定内発話へと誘導するためのヘルプの提示が可能となる．また，当該ドメインに対話制御を移すことで，以後の対話管理を円滑に行うことができる．ドメインとトピックの関係及びその具体例を図 1 に示す．

なお，どのドメインにも共通して含まれるコマンド発話は command-utterance とする．例えば,「そうです」,「検索結果を読み上げて」などがこれにあたる．

3. トピック推定に基づく発話の誘導

本研究での処理の流れを図 2 に示す．まず，入力発話の音声認識を，システムの言語理解に用いる音声認識器

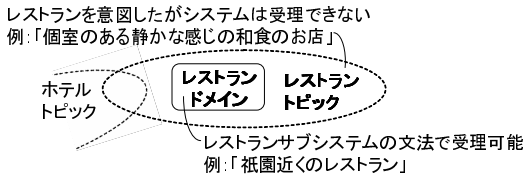

図 1: ドメインとトピックの関係及びその具体例

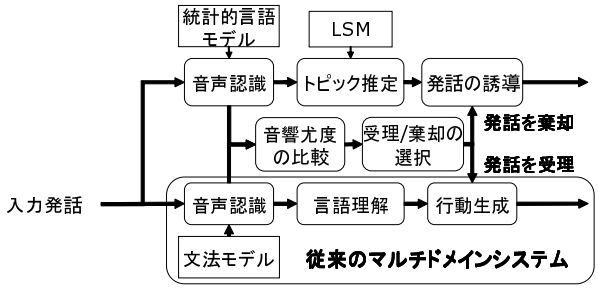

図 2: システムの概略

に加えて，トピック推定のための音声認識器でも並行して行う．次に，２つの音声認識結果の音響尤度を比較し，ある閾値以上なら入力発話の言語理解結果を受理，それ以外は棄却，の取捨選択を行う [3]．受理と判定された場合，言語理解結果に応じてタスクを進行する．棄却の場合は本手法により推定するトピックに応じて対話の制御や発話の誘導を行う．

3.1 LSM を用いた発話のトピック推定

各トピックに対する学習文書集合と入力発話との近さを計算することで，トピック推定を行う．本研究では入力発話と文書集合との近さの計算に LSM を用いる．LSM を用いたトピック推定手法を以下に示す．

学習時 まず，各学習文書に対する単語の頻度をもとに得られる $M \times N$ 共起行列を求める．ここで，M は学習文書集合に現れる異なり単語数，N は学習文書数である．また，推定の対象とするトピック数を n，トピックごとの学習文書数を d とすると，$N = n \times d$ と表される．

その共起行列に対して特異値分解と次元縮約を行い，共起行列の階数を k に減じる．また特異値分解をもとに，N 個の学習文書それぞれに対して k 次元空間でのベクトル表現を得る．

本研究で作成した共起行列は，$M = 67533$，$N = 120$，$n = 6$，$d = 20$ である．レストラン，観光案内，バス，ホテル，天気トピックに関しては，システムの言語理解用文法から生成した文の集合と，Web から収集 [4] した文の集合を d 個に分割し，学習文書を構成した．Web からは各トピックにつき 10 万文を収集し，システムの言語理解用文法からは各トピックにつき 1 万文を生成した．また，command-utterance の学習データとして 175 文を人手で準備した．次元縮約に関しては $k = 40$ とした．

Guiding User’s Utterances based on Topic Estimation for Out-of-Grammar Utterances in Multi-Domain Spoken Dialogue Systems: *Satoshi Ikeda , Kazunori Komatani , Tetsuya Ogata , and Hiroshi G. Okuno (Kyoto Univ.)*

実行時　入力発話の音声認識には，システムの言語理解に用いる文法モデルより広範な統計的言語モデル（トピック推定用言語モデル）を用いる．その音声認識結果に対して，単語の頻度をもとに M 次元ベクトルを求める．それを特異値分解の際に算出した行列を用いて k 次元ベクトルとして，学習時に求めた N 個の k 次元ベクトルとのコサイン距離を計算することで入力発話と学習文書の近さを計算する．ここで，トピックに属する d 個の学習文書と入力発話とのコサイン距離の最大値を，トピックと入力発話の近さと定義する．これにより，入力発話に最も近いトピックを求める．

3.2　トピックが不明な発話の処理

入力発話の中には，トピックが不明な場合がある．トピックがその発話のみでは定まらない文脈依存発話と，システムの対象とするどのトピックにも属さない発話がそれにあたる．例えば,「上限予算五千円」という発話は,その発話のみではホテルトピックとレストラントピックの両方の可能性があり，文脈によってトピックが変わる．また，ホテルやレストランの検索システムに対する「近くの銀行を教えて」という発話は，システムが持つどのトピックにも該当しない．これらのトピックが不明な発話を unknown-topic とし，以下の手順に従って推定する．

まず，入力発話に最も近いトピック T を求める．次に，トピック T の信頼度を $CM_T = closeness_T / \sum_t closeness_t$ として CM_T を求める．ここで，$closeness_t$ はトピック t と入力発話の近さである．$CM_T > \theta_1$ ならトピックの推定結果を T とする．それ以外の場合，トピックの推定結果を unknown-topic とする．unknown-topic は，文脈などの情報をもとに，対話管理や発話の誘導を行う．

3.3　トピックに基づく発話の誘導

本研究の目指す対話と従来の対話の例を図 3 に示す．従来の対話では，システム想定外発話に起因する誤った言語理解をそのまま受理し，応答も誤ったドメインで行ってしまう（図 3 の S-1）．このような場合，ユーザはシステムの誤動作の原因がわからず，どのように言い直せばいいのかもわからない．一方，本研究の目指す対話では，トピック推定を行うことで，内容語の認識はできないが，誤動作することなくユーザの意図を推定し，これに応じたヘルプを提示可能となる（図 3 の S-2）．

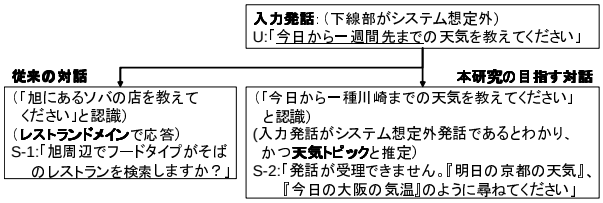

図 3: トピック推定結果を用いた対話の例

4.　評価実験

4.1　評価用データの収集

WOZ 法により 8 名の被験者から対話データを収集した．被験者はレストラン検索，ホテル検索，寺社案内，天気案内，バス運行情報案内の５つの話題を含むシナリオに基づき，自由に対話を行った．この結果，272 発話を得た．収集した発話のうち，発話断片や，書き起こしに内容語が含まれない発話は，システムの言語理解部で棄却されるものであり，トピック推定の対象外であるため取り除いた．正解ラベルはレストラン，ホテル，観光案内，バス，天気，command-utterance，unknown-topic のいずれかを人手で与えた．

表 1: 各手法におけるトピック推定の正解率

	正解率
従来手法	25.0% (68/272 , $\theta_2 = 0.80$)
LSM によるトピック推定手法	59.5% (162/272 , $\theta_1 = 0.26$)

4.2　トピック推定の評価

以下を従来手法として，本手法とトピック推定の精度を比較評価した．

従来手法　各ドメイン文法に対する音声認識結果の，文としての事後確率に基づきトピックを決定する．具体的には，各ドメイン文法による音声認識結果の文全体の音響尤度が最大のドメインを D とし，$e^{\alpha \cdot score_D} / \sum_d e^{\alpha \cdot score_d} > \theta_2$ なら D に対応するトピックを出力する．それ以外なら unknown-topic とする．$score_d$ はドメイン d の文法による音声認識結果の音響尤度である．α はスムージング係数で，$\alpha = 0.05$ とした．

トピック推定用認識器による音声認識には Julius [5] を用い，単語正解率は 69.6%であった．

音声認識結果に対する，従来手法と本手法の正解率を表 1 に示す．評価データにシステム想定外発話が多く含まれるため，従来手法のトピック推定の正解率は 25.0%と低い値になっている．これに対して提案手法では，音声認識結果が 7 割程度であるにもかかわらず，34.5 ポイント高い精度でトピックを正しく推定できる．これによりトピックに応じたヘルプの生成や対話制御ができると期待される．

5.　おわりに

本研究では，ユーザの多様な発話に対する適切な対話管理を目的として，LSM を用いたトピック推定について報告した．被験者 8 名の発話データによる評価実験では，本手法により従来手法に比べ 34.5 ポイント高い精度でトピック推定が行えることを確認した．今回はトピック推定にのみ焦点をあてたが，今後はこれをマルチドメイン音声対話システムへと実装し，トピック推定を利用した対話管理の有効性を評価する予定である．

謝辞　本研究は科研費，21 世紀 COE プログラム，SCAT の支援を受けた．LSM の学習データの収集には京都大学河原研究室で開発された Webcollect[4] を用いた．関係各位に感謝する．

参考文献

[1] 神田，駒谷，中野，中臺，辻野，尾形，奥乃. 複数ドメイン音声対話システムにおける対話履歴を利用したドメイン選択の高精度化. 情処第 68 回全大，5M-1，2006.

[2] J. R. Bellegarda. Latent semantic mapping. *IEEE Signal Processing Mag.*, vol. 22, no. 5, pp. 70-80, Sept. 2005.

[3] 福林, 駒谷, 尾形, 奥乃．音声対話システムにおける発話検証を利用したシステム想定外発話の誤受理抑制，情処第 69 回全大 (2007) .

[4] T. Misu, T. Kawahara. A bootstrapping approach for developing language model of new spoken dialogue systems by selecting Web texts. In *Proc. Interspeech*, pp. 9-12, 2006.

[5] 河原，李．連続音声認識ソフトウェア Julius．人工知能学会誌，Vol. 20，No. 1，pp. 41-49, 2005.