音声対話処理のための発話単位のトピック推定

浅見 克志　　竹澤 寿幸　　菊井 玄一郎

ATR 音声言語コミュニケーション研究所
京都府相楽郡精華町光台 2-2-2
{katsushi.asami, toshiyuki.takezawa, genichiro.kikui}@atr.co.jp

本稿では，音声対話処理のための，発話毎のトピック推定手法について述べる．音声対話処理におけるトピック推定には，満足すべき要件として(1)短い文に対応，(2)幅広いトピックに対応，(3)音声認識誤りに対するロバスト性，などが求められる．本稿では，これらの観点からトピック推定手法を提案し評価する．提案手法は，単語とトピックのエントロピー的観点による関連度を利用する．単語毎にトピックとの関連度を求めていること，反復処理がなく処理負荷が軽いことが，(1)(2)の要件に対してそれぞれ効果的である．また，(3)の要件に対しては，読み・品詞が同じ単語をマージした単語の取り扱いが効果的である．実験により提案手法の性能を検証した．

TOPIC DETECTION OF AN UTTERANCE FOR SPEECH DIALOGUE PROCESSING

Katsushi Asami, Toshiyuki Takezawa, Genichiro Kikui

ATR Spoken Language Translation Research Laboratories
2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, Japan

This paper proposes a topic detection method suitable for speech dialogue processing. There are three requirements for realizing such topic detection. The first is to obtain information from a short utterance. The second is to deal with a broad topic. The third is to have robustness against speech recognition errors. Our topic detection method provides solution to these requirements. A relation factor between the topics and the words, and its generation process without iteration are effective on the first two requirements. On the third requirement, merging words with different surface forms is effective. We verified the performance of the method by experiments using travel conversation corpus.

1. はじめに

トピック推定に関しては，書き言葉では新聞記事などの文書分類[1]，音声言語ではニュース音声の分類[2]～[4]，さらにより日常的な音声言語では，自然言語による call steering[5]～[7]などがあり，これらは広く研究されている．これらの分野においてトピック推定は，情報検索に対する効果や，コールセンタで，発話されたユーザ要求に対応する部門への自動通話振り分け等の利用が期待されている．一方，本稿で提案するトピック推定手法は，音声対話 HMI や音声翻訳への利用を念頭においている．音声対話 HMI ではユーザの要求推定，また，それに応じた対話スクリプトの選択・切換が期待できる．音声翻訳ではトピックに応じた訳し分け・絞り込みが考えられる．さらに，音声認識に対して，話題に応じた言語モデルの自動選択という寄与も考えられる．

トピック推定を音声対話 HMI や音声翻訳等の音声対話処理に利用するには，次のような課題に対処できなければならないと考える．

(1) 幅広いトピックを扱う．
(2) 話し言葉特有の，文長の短い発話に対応する．

表 1 トピックタグの例

Layer 0	Layer 1	Layer 2	例
ACTIVITY	Shopping	Choose something Buy something …	これを見せてくれますか 日本円で買い物ができますか お支払いはどうなさいますか
	Sightseeing	Enjoy sport Take a picture	野球場がどこにあるか知っているの 写真を撮っていただけませんか
TRANSPORT	Move	Use a rental car Buy a ticket …	燃費のいいものはありますか ミネアポリス行きの列車に乗るにはこの ホームでよいのですか
	Airport	Go through immigration Transfer	パスポートと入国カードを見せてください ジェイエル 739 便に乗り継ぐのですが
	Airplane	Use the onboard services Have the onboard meal	映画はどのチャンネルですか どんな飲み物がありますか
10 分類	20 分類	252 分類	←分類数

(3) 1発話毎に推定を行なう. [1]

(4) 音声認識誤りに対してロバストである.

本稿ではトピック推定の対象として, 話し言葉の音声翻訳で広く研究されている旅行対話を取り上げた. 旅行対話で出現するトピックは搭乗手続き, 宿泊, レストラン, 買い物, トラブルへの対応など, 多岐に渡っており, この点では新聞記事やニュース等の特徴と共通する. 一方で, 話し言葉の特徴を有することは, call steering にも共通する. このことから, 旅行対話におけるトピック推定は興味深い. 新聞記事やニュースは, 数十語以上の単語を含んでいると考えられるが, 旅行対話は, 話し言葉であるために, 1 発話あたりの語数は平均で 10 語以下[2]である. この点で, 新聞記事やニュース音声に比べ, トピック毎にはっきりした特徴の抽出が難しくなると考えられる.

以降, まず **2** でトピック情報付きのコーパスである「大規模旅行対話表現集」を簡単に紹介する. このコーパスは, 旅行中に遭遇する様々な場面で現れ得る話し言葉表現を集めたものである. 表現集形式であり, 人間－人間の一連の対話形式になっているものではない. **3** では単語とトピックの関連度と, それを利用して入力発話テキストに含まれる単語からトピックを推定する機構について述べる. この関連度は, 訓練データを基に, 単語とトピックの相互情報量と, 単語のエントロピーから求められる. **4** では, 提案手法によるトピック推定の実験結果を示す. この実験では, 分類問題に対して優れた性能を持つとされるサポートベクタマシン(SVM)を利用する場合, および単語とトピックの相互情報量のみを利用する場合のトピック推定結果と比較する. また, 形態素プロパティの選択と, 音声認識誤りに対するロバスト性についても検討する. さらに推定に使用する単語を品詞によって選別するフィルタの有無に関しても比較する. **5** では, 実験結果に関して考察する. また, 複数の異なる話題間を, ユーザ主導で自由に遷移できる音声対話 HMI システムの例を紹介し, そのような対話システムにおける 1 発話毎のトピック推定の有用性について述べる.

2. 大規模旅行対話表現集

本研究では, 大規模旅行対話表現集[8]を利用して関連度の計算とトピック推定の実験・評価を行なう. このコーパスは, ATR 音声言語コミュニケーション研究所がまとめたもので, 一般のフレーズブック(旅行対話表現集)に見られる表現の日本語・英語の対訳データ集になっており, 約 20 万文から成る. 本研究では, 日本語文のみを使用している.

各発話にはトピック情報が付与されている. このトピック情報は, 一般的なフレーズブックの構成に基づいて付与されている. 通常, フレーズブックは, 利便性を考慮して空港, 航空機内, 出入国審査など場面に分けて構成されているが, 場面分類の表記は, フレーズブックにより異なる. 大規模旅行対話表現集では, これらの様々な分類体系を基に, Layer 0～2 の 3 階層で 10～252 種類のトピックを表すタグを設定した. コーパスに出現する各発話に対して, 3

[1] 音声対話 HMI や音声翻訳が使用される場面を想定すると, 文脈が利用できない状況も有り得る. (**5.2** 参照)

[2] 本稿で使用した大規模旅行対話表現集のテストセットに含まれる発話の平均値

階層それぞれのタグが付与されている. この分類例を**表 1** に示す. なお, 本稿のトピック推定手法では, Layer 0 (10 分類)および Layer 1 (20 分類)について推定する.

なお, このコーパスには「かしこまりました」「どうぞ」など特に応答文で, どのようなトピックにも現れ得る文が含まれる. そのような文に対しては, 対応する質問文・依頼文などのトピックに基づき, 便宜的に, 1回の出現につき, トピックを1つに決めた. ただし, テストセットに関しては, 発話文毎に可能性のあるトピックを全て列挙している.

3. トピック推定手法

本稿で述べる手法では, 入力は形態素情報が付与された単語系列, 出力は Layer 0 および Layer 1 のトピックを示すタグである. 本章では, 本稿のトピック推定手法を訓練フェーズと実行フェーズに分けて述べる.

(i) 訓練フェーズ
訓練データに含まれる全ての単語とトピックの組み合せについて, 関連度を求めておく.

(ii) 実行フェーズ
訓練フェーズで求めた関連度を利用して, 入力発話のトピックを推定する.

3.1 関連度　－訓練フェーズ

関連度とは, 単語とトピックの関連の強さを示す数値である.

本稿で扱うような単語とトピックの関連性の尺度としては, 相互情報量が用いられることが多い. すなわち, ある特定のトピックの出現に依存して出現する単語は, そのトピックとの相互情報量が大きくなり, 逆にトピックの出現と独立に出現する単語は相互情報量が小さくなる性質を利用する[4][9]. しかし, 本研究で使用したコーパスに対しては, どの特定のトピックとも関係がないと考えられる単語の相互情報量が, その単語のエントロピーの増大とともに大きくなる例が, 予備実験で散見された. さらに, その相互情報量は, 特定のトピックに集中して出現する単語の相互情報量を上回った.

実際, **4.2** で示すとおり, 相互情報量を関連度として使用した場合, 良好なトピック推定結果は得られなかった. この問題に対処するため, 単語とトピックの関連性を示す数値として, 相互情報量を補正する必要がある.

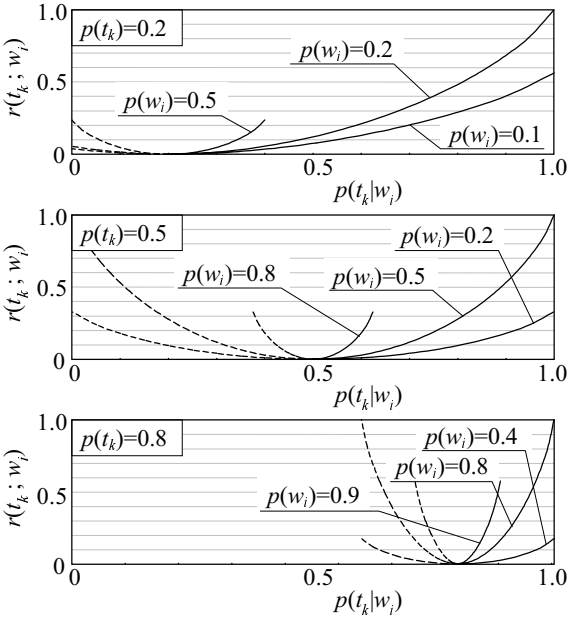

図 1　単語, トピックの出現確率と関連度

ここで,「単語 w_i が出現する／しない」という情報源 W と「トピック t_k が出現する／しない」という情報源 T を考える. 情報源 W のエントロピーを $H(W)$, 情報源 T, W の相互情報量を $I(T;W)$ とする. 発話に含まれる単語から, その発話のトピックを推定するにあたって望まれるのは, 単語側から見たトピックとの関連の強さを示す数値である. それは, 相互情報量と単語のエントロピーの比で与えられると考えられる. そこで, 関連度 $r(t_k;w_i)$ を次のように定義する.

$$r(t_k;w_i)=\begin{cases}\dfrac{I(T;W)}{H(W)} & \left(\dfrac{p_{TW}(t_k,w_i)}{p_W(w_i)}\geq p_T(t_k)\right)\\ 0 & \left(\dfrac{p_{TW}(t_k,w_i)}{p_W(w_i)}<p_T(t_k)\right)\end{cases} \quad (1)$$

$p_T(t_k)$, $p_W(w_i)$ は, それぞれ単語 w_i とトピック t_k の出現確率, $p_{TW}(t_k, w_i)$ は単語 w_i とトピック t_k の同時出現確率である. 式(1)は, 単語 w_i から見たトピック t_k との関連度を示し, $r(t_k;w_i)$ の値の範囲は[0,1]である. **図 1** に単語, トピックの出現確率と関連度の関係を示す. $p_{TW}(t_k,w_i)=p_T(t_k)p_W(w_i)$ (すなわち $p_{T|W}(t_k|w_i)=p_T(t_k)$) つまり T, W が独立ならば, T, W に関連性は無く, $r(t_k; w_i)=0$ となり, $p_{TW}(t_k,w_i)=p_T(t_k)=p_W(w_i)$ (すなわち $p_T(t_k)=p_W(w_i)$ かつ $p_{T|W}(t_k|w_i)=1$) のとき, T, W は完全に関連(一致)しており, $r(t_k;w_i)=1$ となる.

式(1)の，相互情報量のエントロピーによる正規化は，2 つの情報源の情報エントロピー的観点による相関係数である[10]．また，$p_W(w_i)>0.5$ となることは，通常，考え難く，$H(W)$は単語の出現頻度の増加と共に大きくなる．このことから，この正規化は，高頻度語の除去に似た効果があると考えられる．

ところで，式(1)の条件部は，$I(T;W)$を $H(W)$で正規化した値を，次節で述べるトピック推定処理に適用するために付加したものである．この条件を加えない場合，$r(t_k;w_i)$は $p_{T|W}(t_k|w_i)=p_T(t_k)$を中心に対称となり，$p_{T|W}(t_k|w_i)<p_T(t_k)$の領域では「単語 w_i が出現しない場合に，トピック t_k が出現する」関連度も表現している．次節に述べる実行フェーズでは，文中の出現単語に関して関連度を加算するため，単語 w_i が出現しない場合の関連度は不要である．式(1)の条件部により，この領域の関連度を 0 としている．

2 で，トピック推定の対象は Layer 0 (10 カテゴリ), Layer 1 (20 カテゴリ)としたが，関連度は上位階層を全体空間として求める．つまり，Layer 0 では訓練データの全発話が全体空間となるが，Layer 1 では上位階層の Layer 0 で同じトピックタグが付与された発話の集合が全体空間となる．例えば、Layer 1 のトピック“Airport”の関連度を求める場合は，対応する Layer 0 のトピックタグとして“TRANSPORT”が付与された発話の集合を全体空間とする（表 1 参照）．

3.2 トピック推定　－実行フェーズ

トピック推定の基本的な考え方は，各トピックに対する入力文のスコアを求め，最大スコアのトピックを選択することである．トピックに対する入力文のスコアは，入力文に含まれる単語の，トピックとの関連度の総和として求められる．

トピック推定の実行フェーズは，式(2)から式(4)に示す，単純な行列演算が主になる．

Z を入力文 S の特徴を示すベクトルとする．Z の要素は，入力文 S に単語 w_i が含まれる／含まれないを 1/0 で表す．

$$Z=\left[\mu_Z(w_1)\ \mu_Z(w_2)\cdots\ \mu_Z(w_m)\right]$$
$$\mu_Z(w_i)=\begin{cases}1 & : S \text{ includes a word } w_i\\ 0 & : S \text{ does not include a word } w_i\end{cases} \quad (2)$$
$$(i=1,2,\cdots,m)$$

行列 R はトピックと単語の関連度を要素とする．

$$R=\begin{bmatrix} r(t_1;w_1) & \cdots & r(t_n;w_1)\\ \vdots & \ddots & \vdots\\ r(t_1;w_m) & \cdots & r(t_n;w_m)\end{bmatrix} \quad (3)$$

文－トピック関連度ベクトル A はベクトル Z と行列 R の積により求められる．

$$\begin{aligned}A&=Z\cdot R\\ &=\left[\mu_Z(w_1)\cdots\ \mu_Z(w_m)\right]\begin{bmatrix} r(t_1;w_1) & \cdots & r(t_n;w_1)\\ \vdots & \ddots & \vdots\\ r(t_1;w_m) & \cdots & r(t_n;w_m)\end{bmatrix} \quad (4)\\ &=\left[r_A(t_1)\cdots\ r_A(t_n)\right]\end{aligned}$$

トピック推定は Layer 0, Layer 1 について行なうので，実際には n=10(分類; Layer 0)+20(分類; Layer 1)である．

次に下位階層である Layer 1 の文－トピック関連度に，それぞれ対応する Layer 0 の文－トピック関連度を乗算する．例えば，Layer 1 の“Move”, “Airport”, “Airplane”に関する文－トピック関連度には，Layer 0 の“TRANSPORT”の文－トピック関連度が乗算される．この操作により，推定対象の最下位階層のトピックに注目した推定結果ベクトル A_{LOWEST} を得る．ここで，l は最下位階層のトピックの個数である．

$$A_{LOWEST}=\left[r_{A_{LOWEST}}(t_1)\cdots\ r_{A_{LOWEST}}(t_l)\right] \quad (5)$$

最下位階層のトピックが決まれば，それに対応する上位階層のトピックは決まるので，この推定結果ベクトルにより全階層の推定結果が得られる．

1best の推定結果を求める場合は次のように求められる．

$$\hat{t}=\underset{t_k}{\operatorname{argmax}}\ r_{A_{LOWEST}}(t_k) \qquad (k=1,2,\cdots,l) \quad (6)$$

表 2　単語識別例

	表記形	読み	正規形	品詞
A	乗り換え	ノリカエ	乗り換える	動詞
B			乗り換え	名詞
C	乗換		乗換	

“表記形”に注目　: A, B は同じ単語
“読み”に注目　: A, B, C は同じ単語
“正規形”に注目　: A, B, C は互いに異なる
“品詞”に注目　: B, C は同じ単語

3.3 単語識別に使用する形態素プロパティ

単語は“表記形”, “読み”, “正規形”, “品詞”, “活用形”などのプロパティを有する. 単語の識別に使用するプロパティの組み合せによって, 音声認識誤りに対するトピック推定のロバスト性に影響を与える. この影響に関しては次章で述べる. **表 2** に単語識別例を示す.

4. 実験

提案手法の性能を評価する実験を行なった. 実験条件を**表 3** に示す. ここで, テストセット(1), (2)の違いについて説明しておく.

2 で述べたように, 大規模旅行対話表現集では, 1発話の1回の出現について1個のトピックが与えられている. テストセットに対しては, これとは別に, 1人の作業者が判断した, 各発話で可能性のある複数のトピックも与えられている.

テストセット(1)では, 各発話に対するトピック推定結果が, 作業者により列挙されたトピックの何れかに一致する場合, 推定結果を正解として, トピック一致率を求めた.

テストセット(2)は, テストセット(1)から, 作業者による判断の結果, トピックが1個に限定され, かつそのトピックが元来付与されていたトピックと一致する発話のみを抽出したセットである. トピックに関する曖昧性が小さい発話のセットと言える.

なお, この作業者は英日方向版[3]の同様の表現集データ作成に従事しており, トピック分類および発話とトピックの関係について熟知している.

4.1 実験 1

実験 1 では, 提案手法によるトピック推定の結果と, SVM によるトピック推定の結果を比較する. 入力は, テストセット(1)の正解テキストおよび音声認識結果である.

形態素プロパティの組み合せパターンは以下の3種類とした.

(*a*) 全プロパティ
(*b*) 読み＋品詞
(*c*) 表記形＋品詞

[3] 日本を訪れる外国人旅行者向けフレーズブックに基づく旅行対話表現集である.

表 3 実験条件

	訓練セット	テストセット(1)	テストセット(2)
発話数	162320	1523	635
音声認識の WER(%)			12.8

ここで, 提案手法との比較対象とした SVM での訓練データに関して簡単に述べる. まず発話文の特徴を示す属性として名詞類, 動詞類を選択する. 発話文ベクトルは, これらの属性を含む(0)／含まない(1)により, 例えば次のように表現される.

$$\boldsymbol{x}_i = (0, 1, 0, 0, 1, 1, 1, 0, \cdots)$$

入力文が分類対象のトピックに「属する／属さない」は+1/–1 と表現して訓練する.

また, SVM のカーネル関数には以下の多項式関数を使用した[1].

$$K_{poly} = (\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i \cdot \boldsymbol{x}_j + 1)^d \qquad (d = 1) \tag{7}$$

表 4 に実験 1 の結果を示す. 提案手法による推定結果のトピック一致率は, 形態素プロパティのどの組み合せにおいても, また正解テキスト・音声認識結果の何れの入力においても, SVM による場合の一致率を上回った.

次に正解テキスト入力時と, 音声認識結果入力時のトピック一致率を比較する. SVM に関しては, 形態素プロパティの組み合せ条件や Layer によりばらつくが, 正解テキスト入力時に対するトピック一致率の劣化率は, 平均で約7％であり, これは WER に比較して小さい. 一方, 提案手法では, 形態素プロパティ組み合せ(*a*), (*c*)に関して, 劣化率が 11～13%で, WER と同程度の数値となった. これに対し, 形態素プロパティ組み合せ(*b*)では, 劣化率が約 4～7%で, WER に比べて小さくなる結果が得られた.

4.2 実験 2

実験 2 では, テストセット(2)を用いて提案手法によるトピック推定を行なった. 形態素プロパティ組み合せは, (*a*)および(*b*)とした. 実験 1 での結果が(*a*)とほぼ同等であった(*c*)を省いた.

テストセット(1)では, 複数トピックが列挙された発話も存在する. 推定結果が列挙されたトピックの何れか 1 個に一致すれば正解とカウントするので, トピックが曖昧な発話が一致率を押し上げる可能性がある. そこで, 曖昧性が小さいと考えられるテストセッ

表 4　テストセット(1)の発話に付与されたトピックと推定トピックの一致率

トピック一致率(%)	(*a*) 全プロパティ		(*b*) 読み＋品詞		(*c*) 表記形＋品詞	
正解テキスト入力 (WER 0%)	Layer 0	Layer 1	Layer 0	Layer 1	Layer 0	Layer 1
提案手法	80.2	73.0	79.3	72.2	79.9	72.4
SVM	60.3	51.1	58.2	50.7	58.9	42.9
音声認識結果入力 (WER 12.8%)						
提案手法 (正解テキスト入力に対する劣化率)	71.2 (–11.2)	63.2 (–13.4)	75.6 (–4.7)	67.4 (–6.6)	70.9 (–11.3)	63.0 (–13.0)
SVM (正解テキスト入力に対する劣化率)	56.0 (–7.1)	49.9 (–2.3)	54.0 (–7.2)	46.6 (–8.0)	51.0 (–13.4)	41.7 (–2.8)

ト(2)により，より精度良く推定手法の能力を評価できると考えられる．ただし，テストセット(2)に含まれる発話は，作業者によるタグ付けでも 1 トピックに絞られる発話なので，各トピックの特徴が良く現れている発話であるとも言える．

また，この実験では，品詞フィルタにより，content word として名詞類，動詞類，形容詞類，副詞類を選別し，それらを使用して推定を行ない，フィルタの有無の影響を比較した．

図 2 に実験結果を示す．テキスト入力・品詞フィルタなしの条件で，テストセット(1)と比較して Layer 0 のトピック一致率は 2〜3 ポイント上昇した．Layer 1 では，4〜6 ポイント上昇した．また，音声認識結果の入力時では，形態素プロパティ組み合せ(*a*)では，Layer 0, Layer 1 ともテストセット(1)と比較して約 1〜2 ポイント低下した．一方，組み合せ(*b*)では，約 1 ポイント上昇した．さらに，テキスト入力と音声認識入力を比較すると，テストセット(2)の場合でもテストセット(1)と同様，形態素組み合せ(*a*)でトピック一致率の落ち込みが大きい．

これらから，テストセット(1)よりも(2)で推定結果は若干向上するが，傾向に概ね大差ない．品詞フィルタは推定結果には殆ど影響しない．

図 3 に，図 2 の場合と同じ条件で，関連度を相互情報量とした場合の実験結果を示す．まず，提案手法による場合（図 2）と比較して，トピック一致率は大きく劣る．ただし，テキスト入力と音声認識結果入力の比較では，落ち込み幅は小さい．

また，図 2 と図 3 を比較して大きく異なるのは，品詞フィルタの効果である．提案手法では品詞フィルタの効果は殆どない．一方，相互情報量を利用する場合，大きな効果がある．しかし，品詞フィルタを追加しても，提案手法と比較して，トピック一致率には約 10%の差がある．

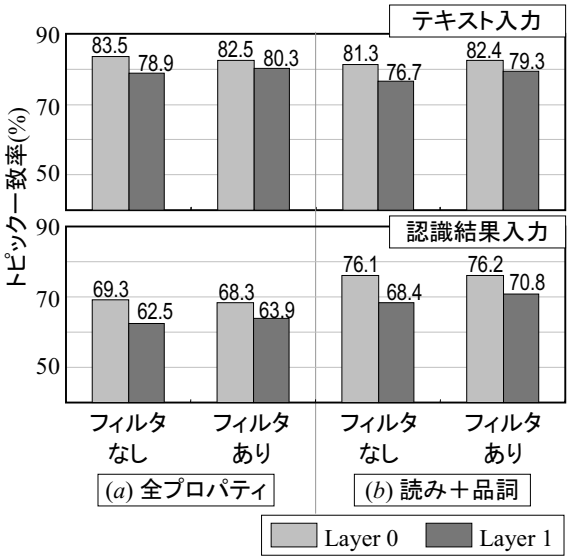

図 2　テストセット(2)でのトピック推定結果

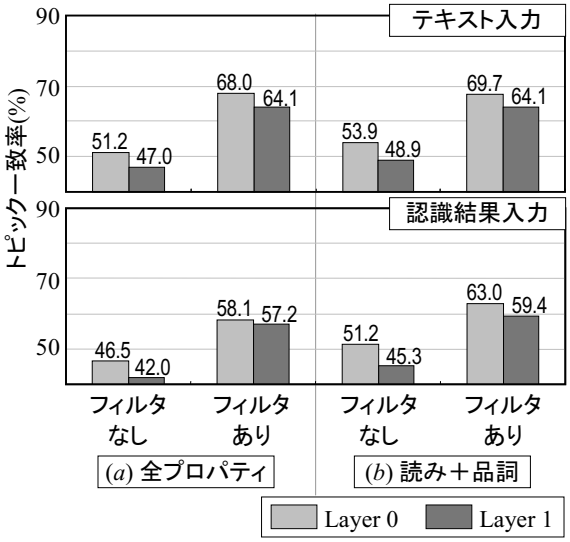

図 3　テストセット(2)でのトピック推定結果
（関連度に相互情報量を利用）

5. 考察

5.1 実験結果に関する考察

まず実験1に関して考察する. 提案手法と SVM の, トピック一致率を比較すると, 提案手法がより高い数値を示しており, 提案手法の方が SVM よりも高いトピック推定能力を持っていると考えられる. SVM は新聞記事やニュースなど多くの語を含み, 多くの属性から topic を推定する場合には有効であるが, 話し言葉の個々の発話に含まれる語数は少なく, 効果的に機能しないと考えられる. SVM を用いる場合は, 単語の意味属性を考慮するなど, 属性を増やすことが必要と考えられる.

音声認識誤りに対するロバスト性に関しては, SVM は WER を基準に考えると, 音声認識誤りの影響を受け難いように見えるが, トピック一致率の絶対値が低いことに注意しなければいけない. 一方, 提案手法は, 形態素プロパティの選択によってロバスト性が変化し, プロパティ組み合せ(*b*)の場合に高いロバスト性を発揮している. プロパティ組み合せ(*a*), (*c*)の場合, 訓練データを詳細にモデル化するため, 音声認識結果に含まれる認識誤りに敏感に反応していると考えられる. プロパティ組み合せ(*b*)は, 訓練データのモデル化が元々粗いため, 多少の音声認識誤りには影響を受けないと考えられる. また, 特定のトピックと関連する語が, 1発話中に2～3語含まれている文が多いことも影響していると考えられる. 例えば

「席を予約したい」(トピック“TRANSPORT”)

という発話では,「席」と「予約」がトピックとの関連度の高い単語である. ここで,

「咳を予約したい(セキヲヨヤクシタイ)」

と誤認識してしまった場合, プロパティ組み合せ(*a*), (*c*)を用いると,「咳」がトピック“TROUBLE”と極めて高い関連度を持ち「予約」の効果を打ち消すため, 誤ったトピックに推定される. 一方, プロパティ組み合せ(*b*)では,「セキ」がトピック“TRANSPORT” “TROUBLE”と,「ヨヤク」がトピック“TRANSPORT” “ACTIVITY”とそれぞれある程度の関連度を持ち, 文全体として“TRANSPORT”と推定される.

次に実験 2 に関して考察する. テストセット(2)では, 正解トピックが1個に限定されるため, 推定結果

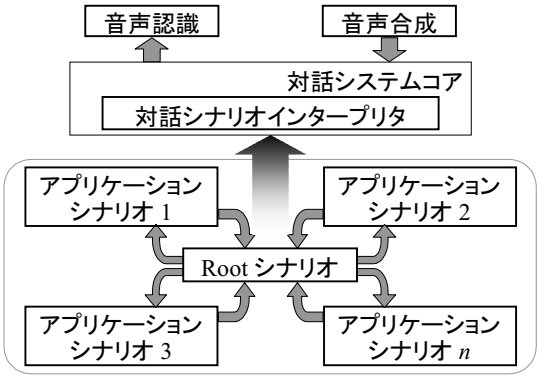

図 4 対話システム例[11]

とのトピック一致率は低下すると予想された. しかし, 実際には, 実験 1 に比べ僅かではあるが, トピック一致率は向上した. **4.2** で述べたテストセット(2)の性質を踏まえると, この結果は推定手法の評価として信頼できると考えられる.

また, 形態素プロパティ組み合せ(*b*)の, 音声認識誤りに対する効果的が, 実験 2 でも示されている.

次に, 図 3 の関連度を相互情報量に換えた場合の結果を見ると, 図 2 に比べて一致率が低く, 本稿で扱う問題に対して, 相互情報量が効果的でないことが示されている.

ここで興味深いのは, 品詞フィルタの効果である. 相互情報量を用いる場合, 品詞フィルタは大きな効果を発揮しているが, 提案手法では, フィルタの有無は推定結果に殆ど影響していない. 品詞フィルタは, トピック推定時に content word 以外を無視することを意味している. 対話に現れる文では,「話者の立場[4]」を介して機能語等がトピックと間接的・潜在的に関連している可能性がある. 式(1)で示される関連度は, この影響を抑えていると考えられる. また, 機能語類は高頻度で出現するので, 品詞フィルタの効果の現れ方に関する相違は, 式(1)で示される関連度が所期の効果を発揮していることを表している.

5.2 対話 HMI におけるトピック推定の応用

音声対話 HMI において考えられる, トピック推定の具体的な応用例を挙げておく.

音声対話 HMI では, 複数ドメインの処理が可能なシステムが必要であり, マルチドメイン音声対話システムが提案されている[11][12]. その例を**図 4** に

[4] 例えば, 旅行対話では「宿泊客」と「フロント係」等の, 話者の役割.

示す. ここでのドメインとは, 例えばカーナビゲーションシステムであれば, ルート設定や施設検索, ネット接続による情報検索などの“機能”である. トピック推定の応用を考えるとき, “トピック”を“機能”と置き換える. すなわち, ユーザの発話は, 「どの“機能”を利用したくて発せられたものか？」を推定する.

図4で, アプリケーションシナリオとはそれぞれ特定の機能を利用するためのユーザとの対話戦略を記述するスクリプトであり, Root シナリオとは, アプリケーションシナリオ間の遷移を司るスクリプトである. これらのスクリプトは, 共通のシナリオインタープリタで処理され, ユーザとの対話が実行される. ユーザが何らかの機能を使用中でも, 予期できないタイミングで他の機能の使用を要求する場合も考えられる. 例えば, 施設情報の検索中に突然, 渋滞情報の検索を要求する, 等である. これに対し, このシステムはいつでもシナリオの遷移ができるようになっている. あるアプリケーションシナリオの処理中に, そのシナリオで対処できないユーザ発話が入力されると, その発話は Root シナリオ処理に送られ, Root シナリオでは, その発話の処理に適したアプリケーションシナリオを選択し, そのシナリオの処理に遷移する. このシナリオ選択は, [11]では予め定義した要求語によっているが, この部分に, 1発話毎のトピック推定が可能である本稿の提案手法を応用できる. 網羅性の面では, 事前定義の要求語を用いるより, 提案手法の有用性が高いと考えられる.

6. むすび

本稿では, 単語とトピックの相互情報量と, 単語のエントロピーから求められる, 単語とトピックの関連度を利用して, 入力発話文のトピックを推定する手法を提案した. 実験の結果, 話し言葉による対話に出現するような文に対しては, 他の手法に比べて効果的にトピックを推定できることが示された. また, 形態素プロパティの組み合せにより, 音声認識誤りに対するロバスト性が変化することが示された. とくに「読み＋品詞」の組み合せにおいて, 認識誤りの影響を受けにくい.

幅広いトピックを持つ大規模旅行対話表現集を利用した実験による結果から, 本稿冒頭に述べた音声対話処理のためのトピック推定の要件に対して, 本提案手法は効果的であると考えられる.

本稿では, トピック推定手法を音声対話 HMI などの対話処理に用いる, ひとつの要素技術として捉えている. 今後は, 実際に対話システムにトピック推定を組み込み, その効果を検討する必要がある.

謝辞

本研究は通信・放送機構の研究委託「大規模コーパスベース音声対話翻訳技術の研究開発」により実施したものである.

参考文献

[1] 平 博順, 春野 雅彦: Support Vector Machine によるテキスト分類における属性選択, 情報処理学会論文誌, Vol. 41, No. 4, pp.1113-1123 (2000)

[2] 櫻井 光康, 有木 康雄: キーワードスポッティングによるニュース音声の索引付けと分類, 信学技報, SP96-66, pp.37-44 (1996)

[3] 横井 謙太郎, 河原 達也, 堂下 修司: 単語の共起情報を用いたニュース朗読音声の話題同定機構, 信学技報, SP96-105, pp.71-78 (1997)

[4] 大附 克年, 松岡 達雄, 松永 昭一, 古井 貞熙: ニュース音声を対象とした大語彙連続音声認識と話題抽出, 信学技報, SP97-27, pp.67-74 (1997)

[5] Wright, J.H., Gorin, A.L., Riccardi, G.: *Automatic acquisition of salient grammar fragments for call-type classification*, Eurospeech 97, pp. 1419-1422 (1997)

[6] Gorin, A.L.: *Processing of semantic information in fluently spoken language*, Proc. ICSLP 96, pp. 1001-1004 (1996)

[7] Chou, W., et al.: *Natural language call steering for service applications*, Proc. ICSLP 2000.

[8] Takezawa, T., et al.: *Toward a broad-coverage bilingual corpus for speech translation of travel conversations in the real world*, LREC 2002, pp. 147-152 (2002)

[9] 鷹尾 誠一, 緒方 淳, 有木 康雄: ニュース音声に対する検索方法の比較, 情処研報, SLP-29, pp.97-102 (1999)

[10] 堀部 安一, 情報エントロピー論, 森北出版 (1989)

[11] 笹木 美樹男, 浅見 克志: 車載マルチメディアにおける対話エージェントについて, シンポジウム「カーナビ・携帯電話の利用性と人間工学」, pp.165-170 (2000)

[12] 長森 誠, 河口 信夫, 松原 茂樹, 外山 勝彦, 稲垣 康善: マルチドメイン音声対話システムの構築手法, 情処研報, SLP-31, pp.45-52 (2000)