

雑談中の潜在的話題遷移を考慮したユーザーの意図推定の検討*

佐古淳，滝口哲也，有木康雄 (神戸大)

1 はじめに

近年，音声による機器操作インターフェイスが実用化されつつある．特に，ロボットとのコミュニケーションやカーナビ操作など，手を使うことが困難な機器への適用がなされている．音声を用いる際，発話が機器操作を意図したものであるか，または同乗者との雑談のように機器とは無関係なものであるかを判別する必要がある．このため，物理的なスイッチを用いて判別を行う手法や，ネットワーク文法を用いて自動的に区別を行う手法 [1] が提案されている．また，我々も音声認識結果に対してブースティングを用いることで，システム要求発話の柔軟性・多様性を保持したまま自動的に区別を行う手法の提案を行ってきた [2] ．

一方で，発話には話題があり，話題は時間とともに変化していく．また，システムへの要求の有無とそのときに話されている話題には相関があると考えられる．例えば，昼食についての話題の際にはカーナビへの要求があるかもしれないが，政治や経済の話題の際に要求があることは少なくなるであろう．本研究では，話題の遷移を pLSA[3] に基づく Topic HMM[4] によって追従し，話題を考慮したユーザーの意図推定を行う．

2 提案手法

本研究では，音声から，潜在的な話題の遷移を考慮した意図推定を行う．潜在的な話題の分析手法として pLSA を用い，話題分布の変化を HMM による状態遷移モデルで表現する [4] ．概要を図 1 に示す．雑談や要求を含むユーザーの発話が潜在的な話題の状態に依存し，また話題の状態が遷移していくモデルを仮定した．発話数を N とし，音声の系列を $\mathbf{O} = (\mathbf{o}_1, \dots, \mathbf{o}_N)$ ，文章の系列を $\mathbf{D} = (\mathbf{W}_1, \dots, \mathbf{W}_N)$ ，潜在的な話題の状態系列を $\mathbf{S} = (s_1, \dots, s_N)$ ，意図の系列を $\mathbf{R} = (r_1, \dots, r_N)$ とすると，提案手法は以下のように定式化される．

$$P(\mathbf{D}, \mathbf{S}, \mathbf{R} | \mathbf{O}) \approx \underset{\mathbf{D}, \mathbf{S}, \mathbf{R}}{\operatorname{argmax}} \prod_{i=1}^N P(s_i | s_{i-1}) P(\mathbf{W}_i | s_i) P(r_i | \mathbf{W}_i, s_i) P(\mathbf{o}_i | \mathbf{W}_i).$$

ここで，潜在的な話題の状態 s_i のみ一時刻前の状態 s_{i-1} に依存し，文章 $\mathbf{W}_i$ ，意図 r_i ，音声 $\mathbf{o}_i$ はそれぞ

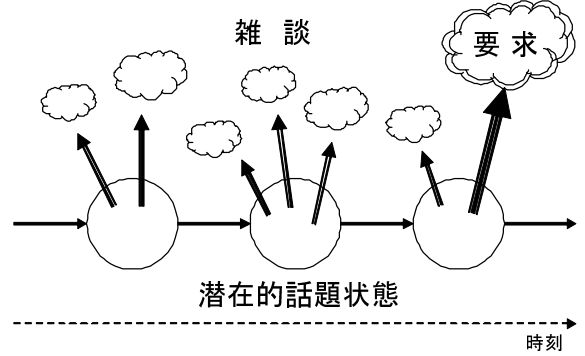


Fig. 1 提案手法の概要

れ同時刻の s_i ， $\mathbf{W}_i$ 及び s_i ， $\mathbf{W}_i$ にのみ依存すると仮定した．

定式化により，右辺第 1 項から 4 項までの 4 つの確率モデルが得られる．第 1 項は，話題の状態遷移確率であり，話題分布の変化をモデル化した Topic HMM の状態遷移確率を用いる．第 2 項は，状態依存の言語モデルである．第 1 項，及び第 2 項については次節において述べる．第 3 項は，話題，及び認識仮説 $\mathbf{W}_i$ からユーザの意図 r_i を推定するモデルである．これは，2.2 節において述べる．第 4 項は，通常の音響モデルと同様である．

2.1 pLSA に基づく Topic HMM

pLSA では，文書 d における単語 w の生起確率 $P(w|d)$ を，潜在変数 z を用いて，

$$P(w|d) = \sum_z P(w|z)P(z|d)$$

のように表現する． z を潜在的な話題と考え，文書における単語生起確率を，潜在的な話題ごとの単語生起確率と文書の話題分布に分解する手法である．ここで， d が時系列を持った文書系列である場合，文書 d の話題分布 $\phi_d = \{P(z|d)\}_{z=1}^Z$ を観測特徴ベクトルとして Ergodic HMM を構築することが出来る [4] ．これにより，話題分布の変化を状態遷移モデルで表現できる．学習された HMM は，状態間の遷移確率と，状態ごとに話題分布の平均ベクトルを持つ．状態遷移確率は，話題の変化しやすさを表しており，定式化の第 1 項に相当する．状態の平均ベクトルは，その状態における典型的な話題分布であると考えられる．定式化の第 2 項である状態依存の言語モデルは，通常の trigram を状態の持つ典型的な話題分布で言語モデ

*Studies on Intent Estimation using Latent Topic Transition in Conversations, by Atsushi SAKO, Tetsuya TAKIGUCHI and Yasuo ARIKI (Kobe University)

ル適応することにより得る．適応手法には Unigram rescaling を用いる．すなわち，状態 s における話題分布 $P(z|s)$ ，及び通常の trigram $P(w_j|w_{j-2}^{j-1})$ を用いて，

$$P(w_j|w_{j-2}^{j-1}, s) = \frac{P(w_j|s)}{P(w_j)} P(w_j|w_{j-2}^{j-1})$$

$$P(w_j|s) = \sum_z P(w|z)P(z|s)$$

により計算する．

2.2 AdaBoost による意図推定

AdaBoost[5] は，単体では識別能力の高くない二値の弱識別器 $h_t(\mathbf{W}, s)$ を複数組み合わせることで識別能力の高い二値の強識別器 $f(\mathbf{W}, s)$ を構成する手法であり，以下のように定式化される．

$$f(\mathbf{W}, s) = \frac{1}{\|\alpha\|_1} \sum_t \alpha_t h_t(\mathbf{W}, s).$$

ここで α_t は弱識別器 h_t の重みである．テキスト分類のための弱識別器としては，単語の有無による手法が提案されている．本研究では，単語に加え，Topic HMM における状態も識別のための情報として用いた．AdaBoost は確率に基づく手法ではないことから，提案手法のモデルとしてそのまま用いることはできない．そこで近似的に sigmoid 関数により値域 0~1 の値に変換して用いる．

$$P(r_i = +1|\mathbf{W}_i, s_i) = \text{sigmoid}\{f(\mathbf{W}_i, s_i)\}$$

$$P(r_i = -1|\mathbf{W}_i, s_i) = 1 - \text{sigmoid}\{f(\mathbf{W}_i, s_i)\}$$

$$\text{sigmoid}(f) = \frac{1}{1 + \exp(-w_1 f - w_0)}.$$

提案手法の定式化第 3 項は上記のモデルを用いる．

3 実験

評価はシステムへの要求検出タスクにおいて行った．これは，人間 2 人とシステムが存在する環境において，雑談を行いつつ任意にシステムに要求を出すものである．システムへの要求を正しく検出できたかについて再現率・適合率・F 値を求めた．

システムとして，音声操作により移動や写真撮影などの動作を行うロボットを用いた．典型的な利用方法として，少し離れた場所から「こっちに来て」とロボットを呼ぶ，「写真を撮って」と写真を撮ってもらう，などがある．研究室において，2 人の人間が会話を行いながら，任意にロボットに対して要求発話を行った．収録はロボットのマイクではなく，2 人の発話者それぞれの胸元に取り付けたマイクで行った．発話数は 330，内 52 発話がシステム要求発話であった．

3.1 音声認識条件

音響モデルは，まず，CSJ モニター版のうち男性話者 200 名の講演音声を用いて作成し，これにテ

ストセット・クローズドな話者適応を行ったものを用いた．適応データの分量は，約 10 分であった．言語モデルは，実験で用いた発話を書き起こしたテキストから作成した．ただし，テストセットに対してオープンとなるように，話者 B の発話のみを用いて話者 A の認識用言語モデルを作成した．音声認識の結果，単語正解精度は 42.1% であった．このような音声認識条件の下でワードグラフを構築し，実験に用いた．

3.2 実験結果

実験の結果を図 2 に示す．Words+Topic が提案手法，Words は潜在的話題を考慮しない従来手法である．提案手法において Precision が向上し，要求検出のわきだしが減少している．しかしながら，効果は限定的であった．理由として，話題に関わらず会話の切れ目の度に，システムへの要求発話を行っていたことが考えられる．より自然な状況においてコーパスを収集すること，及びユーザーの意図・欲求に深く関わったタスクを用いることが望ましい．

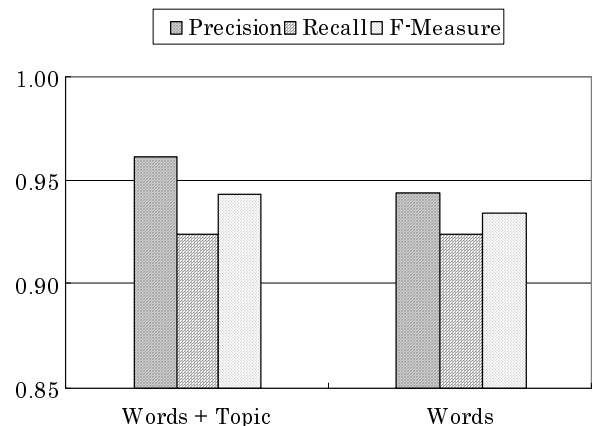


Fig. 2 実験結果

4 おわりに

本稿では，ユーザーの意図推定において，潜在的な話題の遷移を考慮する手法について述べた．評価実験の結果，話題の考慮によりわきだし誤りが減少することを確認した．今後，研究室・ロボットという実験環境ではなく，実際の車内においてカーナビを用いた実験を行う予定である．

参考文献

- [1] 石塚, SP98-5, Apr. 1998.
- [2] 佐古, 音響論(春), pp. 22-23, 2007.
- [3] T. Hofmann, 15th UAI, pp.289-296, 1999.
- [4] A. Sako, Interspeech2007, pp.606-609, 2007-09,
- [5] R.Schapire, Annals of Statistics, vol.26, no.5, pp.1651-1686, Oct. 1998.