使用履歴に基づくユーザー嗜好を考慮した POMDP による音声対話システム*

☆藤川賢至，滝口哲也，有木康雄 (神戸大・工)

1　はじめに

近年，音声対話システムは様々な分野において急速に発展してきた [1, 2]．しかし，発話変形や，発話誤り，意味の曖昧性のために，ユーザー発話の意図を確実に決定することは難しい．その結果誤動作が生じ，音声認識を利用するユーザーが増えない理由となっている [3, 4]．また，システムがユーザーの行為を導く発話をしたとしても，ユーザーの行為は一意には定まらない．現在のインターフェースが未熟であるため，誤認識からの回復も困難となっている．

本稿では，不確定な情報に対しても対話を制御させるために，部分観測マルコフ決定過程（POMDP）を用いる．この手法によって，雑音のある状況下で誤認識が起こった場合でも，自然な対話の中で回復することが可能となる．

また，行動モデルは POMDP において確率的に扱う．これはシステムが環境に与える影響を，必ずしも 100%予測できるとは限らないからである．このモデルに使用履歴から得た嗜好を導入することで利用者の発話予測を行い，誤認識の際も修正できるようにする．本研究ではユーザーシミュレーションによって取得した使用履歴からユーザーの嗜好を得て，行動モデルに反映させることで利用者の意図した発話を扱えるようにした．本稿では，シミュレーション実験を行い，提案手法の有効性を示す．

2　POMDP の概要

一般的に POMDP は $\{S,\ A,\ T,\ O,\ Z,\ R\}$ で表される．$s \in S$ は状態を表す（システムとユーザーの状態）．$a \in A$ はシステム側のアクションを表す．また T はアクション a によって，状態 s が s' へ変わる状態遷移確率 $P(s'|s,a)$ の集合である．$o \in O$ はユーザーや環境から観測される観測値を表し，Z はアクション a によって状態が s' に遷移した後，観測値 o' が観測される観測値出力確率 $P(o'|s',a)$ の集合である．$r(s,a) \in R$ は，状態 s でアクション a を行った時の期待報酬を表す．それぞれの変数の関係をダイナミックベイジアンネットワークで表すと Fig. 1 のようになる．

POMDP は以下のように動作する．まず，システムがアクション a を実行すると，状態 s' へ遷移する．

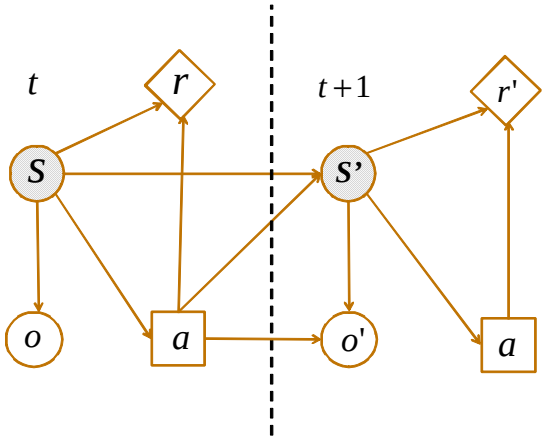

Fig. 1　DBN を用いた POMDP の影響図

この時，s' は s と a のみに依存して決まる．その状態変化により，観測値 o' が観測される．そして，アクション a と状態 s で規定されている報酬 r を得る．これを対話が終了するまで繰り返す．POMDP では状態を直接観測出来ないので，確率分布として扱う．その分布を $b(s)$ とする．この分布 $b(s)$ が分かっている時，アクションによって次の時刻の分布 $b'(s')$ は次式で表される．

$$b'(s') = k \cdot P(o' \mid s', a) \sum_{s \in S} P(s' \mid s, a) b(s) \qquad (1)$$

ここで k は $b'(s')$ の総和を 1 にするための正規化係数である．これを用いると，システムが現在時刻を 1 として，将来時刻 t までに得られる平均報酬は次式で表される．ここで Υ（<1）は割引率を表す．割引率は将来の報酬が現在においてどれだけの価値があるかを決定する．

$$V_t = E[\sum_{\tau=1}^{t} \Upsilon^{(\tau-1)} \sum_{s} b_\tau(s) r(s, a_\tau)] \qquad (2)$$

POMDP の学習では，(2) 式を最大にするような方策を強化学習により求める．方策は，将来獲得できる報酬を最大にするアクション a を時間に独立に b のみから選択出来る．

3　個人嗜好を用いた対話エージェント

誤認識の要因は，発話意味の曖昧性，発話誤りの多さ等に起因する．その結果として，誤動作が生じ，

*Spoken dialogue system using POMDP incorporating user preference based on usage history, by Kenji Fujikawa, Tetsuya Takiguchi, Yasuo Ariki (Kobe University)

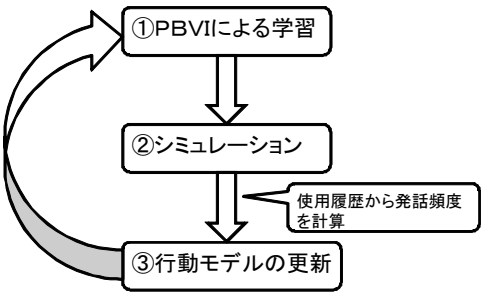

Fig. 2　導入の流れ

音声認識を利用するユーザーに負荷をかける原因となっている．

そこで本稿では，対話制御部に POMDP を導入した．システムはユーザーのゴールを確率分布により求めているので，たとえ誤認識を起こした場合でも，回復が可能である．図 2 で具体的な導入の流れを示す．まず初期設定により与えられた値を用いて PBVI(Point-Based Value Iteration) による学習を行い，POMDP を用いてユーザーシミュレーションを行う．その際，ユーザーが各目的地を何回設定したかという使用履歴を作成する．目的地に設定されればその目的地へ向かう遷移確率を増加させ，それ以外の遷移確率を減少させるといった様に遷移確率の更新を行う．そして更新した値を用いて再学習を行う．これを繰り返すことでユーザーが頻繁に登録する目的地への遷移確率が増加し，ユーザーの嗜好にあった最適な方策を構築することができる．

本稿では，目的地登録を扱うタスクを設定した．ユーザーの行きたい目的地を聞き出し，その目的地を登録するというものである．このシステムの目的はユーザーの発話を誤認識した際でも，再度質問することで正しい目的地を登録することである．システムのアクションはユーザーの発話について，聞き直す，確認を取る，ユーザーのゴールを決定するの 3 種類がある．

4　評価実験

評価実験として，ユーザーシミュレーションによる試行を 500 回繰り返し目的地設定の成功率を調べる．また 500 回のシミュレーションが終わるとその使用履歴から再学習を行い，再びシミュレーションによる試行を行う．

この学習とシミュレーションのサイクルを 3 回行った結果を図 3 に示す．横軸は学習回数を表しており，初期設定の際の学習と，シミュレーションによって得られた履歴を用いて再学習を行った回数の総和を表している．縦軸は 500 回のシミュレーションによる成功率を示す．

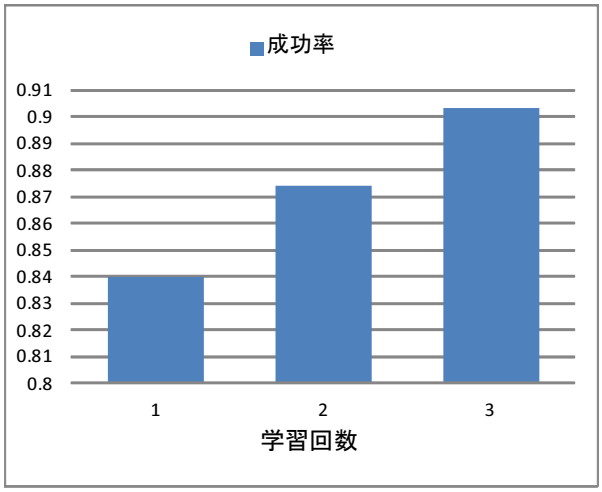

Fig. 3　実験結果

結果から，使用履歴から遷移確率を更新するほど成功率が増加していることが分かる．これはユーザーの嗜好を考慮して行動モデルを更新したことにより，全く予測できなかったユーザーの発話を予測しやすくなったためである．そのため誤認識が起こった際でも，システムが目的地を聞きなおすなどの行動を取ることで，修正することが可能になったと考えられる．

5　おわりに

本稿では，POMDP に使用履歴から嗜好性を導入し，シミュレーションによる実験を行った．その結果，ユーザーの発話を予測しやすくなり，誤認識の際でもシステムが聞きなおすことができるようになり，成功率の増加につながった．しかしながら対話回数が増加してしまうという問題がある．対話をできるだけ少なくし，成功率を上げるために報酬の設定を考える必要がある．今後の課題として，タスク拡大を図るとともに，より誤認識に強く，直観的に使える音声認識システムの実現に取り組んでいく．

参考文献

[1] S. Young,“ The statistical approach to the design of spoken dialogue systems, ”Technical Report CUED/FINFENG/TR.433, Cambridge University Engineering Department, 2002.

[2] P. J. Gurston, et al. “ Oasis natural language call steering trial, ” In Proceedings of INTERSPEECH, pp.1323-1326, 2001.

[3] 西村雅史,“ 音声認識ビジネスの現状と将来展望, ” IPSJ, 2005-SLP-55-3, pp. 13-15, 2005.

[4] 神沼充伸, “ 自動車用音声インターフェースへの期待, ” PSJ, 2006-SLP-63-9-2, pp. 47-48, 2006.