音声言語情報処理 42－2
（2002．7．12）

バス運行情報案内システムにおける ユーザモデルを用いた適応的応答の生成

上野 晋一　　駒谷 和範　　河原 達也　　奥乃 博

京都大学 情報学研究科 知能情報学専攻

〒606-8501 京都市左京区吉田本町

e-mail: ueno@kuis.kyoto-u.ac.jp

あらまし　音声対話システムにおいては、ユーザに応じて適応的に対話の制御・応答生成を行うことが望ましい。本研究では開発中の京都市バス運行情報案内システムに、システムに対する習熟度、タスクドメインに関する知識レベル、性急度の 3 つのユーザモデルを導入し、それに応じた対話制御を検討する。また、音声や言語、対話レベルなどの種々の特徴から、本システムによる実対話データを用いて分類したユーザモデルの判別方法の決定木学習を行った。その結果、判別に有効な特徴が示され、習熟度においては 83.6%の精度での判別が実現できた。

Generation of Cooperative Responses using User Model in Spoken Dialogue System

Shinichi Ueno　　Kazunori Komatani　　Tatsuya Kawahara　　Hiroshi Okuno

Graduate School of Informatics, Kyoto University,

Kyoto 606-8501, Japan

e-mail: ueno@kuis.kyoto-u.ac.jp

Abstract　Our goal is to generate cooperative responses to each user in a spoken dialogue system by choosing an appropriate user model. User models are implemented in Kyoto city bus information system which has been developed at our laboratory. The category of user models are skill level to the system, knowledge level of the task domain and the degree of hastiness. We perform decision tree learning using real dialogue data collected by the system picks up features specific to spoken dialogue systems and semantic attributes. The result shows features appropriate for classification and we get accuracy of 83.6% for classification of skill level.

1 はじめに

音声認識技術の進展にともない、音声対話システムに関する研究が数多く行われている。その単純な形態として、音声認識を用いた自動音声応答システム（IVR）が実用化され、身近なものとなりつつある。その例として、電話を用いてニュースや地域情報を得ることができるボイスポータルシステム [1][2] が挙げられる。しかし、現状のシステムはどのような状況でも画一的な応答を行い、ユーザにとって快適な対話が行われているとは言いがたい。

これに対して、音声対話システムにおいて協調的な対話 [3] を行うための研究がなされている。例えば、混合主導対話の戦略 [4] が挙げられる。その戦略では、システムに慣れたユーザには自由な発話を促し、適宜確認や誘導が行われる。また、情報検索において、[5] ではユーザの発話に必要な情報が含まれていない場合や検索結果が多すぎた場合におけるユーザへの質問や、検索結果が得られなかった場合の代替案の提示方法について述べられている。[6] では対話の各時点での入力された情報や検索状況の説明が有効であることが述べられている。

しかし、どのような応答が協調的であるかは、個々のユーザの知識などの性質により異なると考えられる。例えばユーザの沈黙があっても、単純に対話システムに慣れていないだけなのか、対話している事柄について知識が欠けているのかによって、適切な応答は異なる。また、その場を対処することができても、沈黙があったという事実から、どのようなユーザであるか推測し、その後の対話に生かすことができなければ、再び同じ状況に陥ってしまう恐れがある。そこで、本研究では音声対話システムにユーザモデルを導入することで、ユーザに適応的な応答を生成することを考える。

これまで、ユーザモデルの研究は主として自然言語対話システムにおいて、ユーザの知識に重点がおかれて行われてきた [7]。音声対話システムにおけるユーザモデルの研究には、ユーザの発話の意味内容から判断される知識の深さに応じて応答を変える道案内システム [8] や、場所や時間などの周囲の状況と、性別・予算などのユーザ情報を過去のデータベースと照合してユーザの好みを判断し、レストランを紹介するカーナビゲーションシステム [9] などがある。しかし、これらはタスクドメインに大きく依存してユーザの一面のみについてモデル化を行ったものである。実際の対話においては、種々の面からモデル化を行い、それらを統合して応答を生成する必要がある。また、上記の研究では音声の特徴を利用しておらず、基本的に正しいテキスト入力を前提とした処理になっている。

これに対して音声には、バージイン、応答までの時間など、テキストには含まれない情報が多く存在する。このような特徴は対話をスムーズに進める上で有効であるが、ユーザによって大きく異なると考えられる。このような特徴を反映したユーザモデルを推定し、適応的・協調的に応答を生成することが重要である。そこで、本研究ではシステムに対する習熟度、タスクドメインに関する知識レベル、性急度のユーザモデルを導入し、適応的な対話制御・応答生成の実現を目指す。そして、現在我々が開発中の京都市バス運行情報案内システム [10] で収集されている対話データを用いて、ユーザモデルの判別に有効な特徴と規則について決定木学習を行い、本手法の有効性について検討する。

2 バス運行情報案内システム

京都市バス運行情報案内システム [10] は、ユーザの指定するバスが乗車する停留所の何駅手前まで接近しているかを知らせるシステムである。本システムは電話による利用が可能であり、分刻みで変わるバスの情報をどこからでも手軽に知ることができる。ユーザーは乗車場所、降車場所（これらはバス停名か、そのバス停が最寄となる各所・施設）とバスの系統番号を音声で入力し、バスの接近情報を得る。対話例を図 1、2 に示す。

システムの概略を図 3 に示し、主要部分について以下に説明する。

VWS (Voice Web Server)

VXML（VoiceXML）で指定された対話のスクリプトに沿って、音声認識システムや音声合成システムを駆動する。

音声認識部

プロンプトに応じて指定された文法、語彙に沿って発話を認識する。

対話管理部

VWS からの認識結果（乗車バス停名、系統番号など意味情報のみ）をもとに、ユーザへの応答文を作成する。検索条件がそろえば、Web サーバからバスの接近情報を取得する。

S：乗車駅、降車駅、系統番号をおっしゃってください
U：百万遍から 206 系統に乗ります
S：百万遍から 206 系統に乗るのですか
U：はい
S：206 系統の北大路バスターミナル行きのバスは、
しばらく来ません
京都駅行きのバスは、6 駅手前の高木町を出発
しています

図 1: バス運行案内システムの対話例 1

S：乗車駅、降車駅、系統番号をおっしゃってください
U：四条河原町
S：四条河原町からバスに乗車するのですか
U：はい
S：どこの停留所でバスを降りるのですか
U：嵐山です
S：四条河原町から嵐山までバスに乗るのですか
U：はい
S：11 系統の嵐山、山越中町行きのバスは、
2 駅手前の三条京阪前を出発しています

図 2: バス運行案内システムの対話例 2

VXML 作成部

対話管理部からの応答文、及び認識文法の指定情報に基づいて VXML を動的に生成し、VWS に渡す。

3 ユーザモデルの分類と対話戦略

このシステムに対して、ユーザモデルを導入する。具体的には、図 3 においてユーザ判別部とユーザプロファイルの部分を新たに付け加えた。ユーザ判別部は音声認識部が得た情報をもとにユーザの判別を行う。対話管理部はその判別結果であるユーザプロファイルを対話制御・応答生成に利用する。

音声対話システムにおけるユーザモデルに関する研究 [11] では、我慢強い、従順である、習熟しているなどのユーザを想定している。これらは典型的ユーザであるが、本研究では複数の尺度を用いて包括的にモデル化を行う。具体的には、以下の 3 つの分類を用いる。

- システムに対する習熟度（高い／低い）
- タスクド メインに関する知識レベル（高い／低い／不明）
- 性急度（高い／低い）

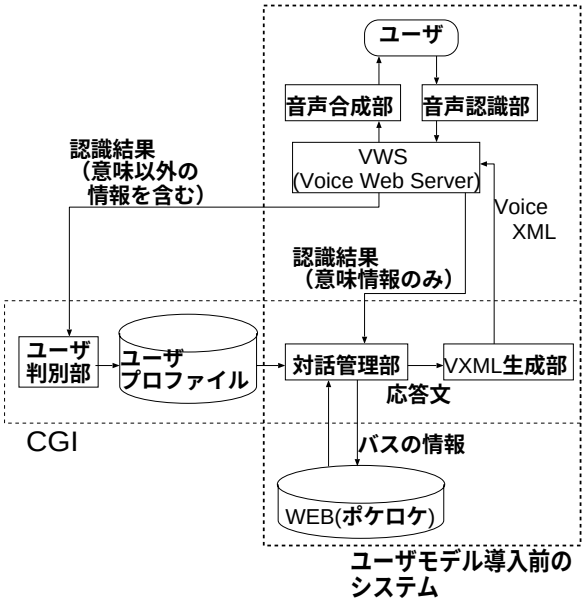

図 3: ユーザモデルのバス案内システムへの導入

以下、それぞれに関する説明と対話戦略について述べる。

3.1 システムに対する習熟度

音声対話システムはまだ一般的であるとはいえず、ユーザには習熟度の差が生じる。習熟度にあわせて、システム主導、ユーザ主導というように応答を切りかえるのが望ましい。従来においては、ユーザが発話しない場合や認識がうまく行かない場合に、システム主導でガイダンスを出す程度であった。判別に有効な特徴として、一回の発話で入力されたスロットの数、バージインの有無、沈黙の有無などが考えられる。習熟度の低いユーザに対しては、単語発話を誘導する質問を行うことや、質問を行う際に答え方の説明を付加するという対話戦略をとる。

3.2 タスクド メインに関する知識レベル

音声対話システムのインタフェースへの習熟度とは別に、当該タスクド メインに関する知識レベルにもユーザ間の差があり、それに対応して提示する情報を変える必要がある。本システムにおいては、京都市は観光地であるので、京都市民と観光客では地域に関する知識に差がある。そこで、タスクド メイ

ンに関する知識というユーザモデルを導入する。バスの運行する地域に関する知識レベルの高いユーザ（京都市民を想定)、低いユーザ（観光客を想定)、および不明に分類して、応答を切りかえる。

判別に有効な特徴としては、認識結果、沈黙の有無が挙げられる。認識結果からは以下を抽出する。

利用するバス停 (キーワード) の属性

バス停は市民のみが利用するバス停とその他のバス停に分ける。バス停の分類は、その最寄りの施設により行う。最寄り施設に寺社、ホテル、駅など観光客が利用する施設がないバス停を市民のみが利用するバス停とする。

場所 (キーワード) の指定方法

乗車場所、降車場所の入力が、正式なバス停名でなされたか、それとも最寄りの施設名でなされたかを、知識レベルの判断材料とする。

知識レベルの低いユーザに対しては、提示するバスの運行情報を少なくして、それについての説明を付加する。一方、知識レベルの高いユーザに対しては、現在のバスの運行情報を詳しく提示する。

3.3 性急度

音声によるコミュニケーションでは、他の手段 (ブラウジング等) に比べて情報提供への切迫性が大きい場合が多いと考えられる。特に本システムでは、携帯電話で移動中に利用される場面が想定されることや、バスの運行情報という分刻みの情報を扱うという性質から、そのような状況が多い。そこで、性急度というユーザモデルを導入し、性急度の高いユーザと低いユーザにあわせて、応答を切りかえる。判別に有効な特徴として、バージインの有無、沈黙の有無、発話継続時間などが考えられる。性急度の高いユーザに対しては最低限の入出力を行うという戦略をとる。

4 決定木によるユーザの判別

ユーザモデルの判別は、ユーザの発話ごとに毎回行う。判別には対話で得られる特徴のみを使用する。ある発話から得られた特徴はユーザ判別部に送られるとともに履歴情報として蓄えられ、ユーザ判別部ではそれらの特徴をもとに判別を行う。本システムで用いる特徴を表 1 に列挙する。これらは、人間が判断に利用していると思われる特徴と、認識スコアなど音声対話システム特有の特徴からなる。

表 1: 判別に用いる特徴

一発話のみに依存する特徴
対話の状態
対話開始からの継続時間
バージインの有無
発話の継続時間
認識の結果 (認識成功・失敗・無入力など)
認識のスコア
入力されたスロットの数
履歴を利用した特徴
発話回数
一回前の対話の状態
現在の質問の連続回数
同じ質問の繰り返される平均回数
対話時間に占めるユーザが発話している時間の割合
全発話に対するバージインのあった発話の割合
一回前の認識結果
全発話に対する発話が認識された割合
全発話に対する認識失敗の割合
全発話に対する発話受理状態になる前に発話が開始された割合
全発話に対する無入力の割合
バージインの回数
発話が認識された回数
認識失敗の回数
発話受理状態になる前に発話が開始された回数
無入力の回数
認識のスコアの平均値
一発話で入力されたスロットの最大数
履歴を利用した特徴 (知識レベルの判別のみに利用)
利用するバス停の属性
場所の指定方法
バス停の属性の割合
場所の指定方法の割合

ユーザの判別の方法として、[12] ではベイジアンネットワークが用いられているが、本研究では複数の特徴から各クラスへの分類に適し、学習が容易であり、学習の結果が直感的にわかりやすいという利点を持つ決定木学習を行う。実対話データをもとに判別のための決定木を学習する。

5 システムの動作例

以上で述べたユーザモデルを実装した。その結果、システムは図 4、図 5 のように応答を行うようになった。図 4 は性急度が高い場合で、入出力をできる限り少なくする戦略をとっている。一方、図 5 は性急度が低い場合で、詳細な入出力を行う。また、習熟度が低いことから、単語発話を誘導する質問を行い、答え方の説明を行う。さらに、知識レベルが低いことから出力情報に説明を加える。

S：利用するバス停、系統番号をどうぞ
U：円山公園から 207 系統
(**習熟度**:**低 性急度**:**高 知識レベル**:**低 の場合**)
S：円山公園から 207 系統でよろしいですか
(**性急度**:**高→確認後、検索開始**)
U：はい
S：九条大宮・九条車庫行きのバスは、
3 駅手前の馬町のバス停を出発しました。
東福寺・九条車庫行きのバスは、
3 駅手前の四条西洞院を出発しました。

図 4: ユーザモデル導入後の動作例 1

S：利用するバス停、系統番号をどうぞ
U：円山公園から 207 系統
(**習熟度**:**低 性急度**:**低 の場合**)
S：どこでバスを降りるのですか
(**習熟度**:**低→単語発話誘導**)
バス停名か、最寄りの有名施設を教えてください
(**習熟度**:**低→説明付加**)
U：東福寺
S：円山公園から東福寺まで 207 系統のバスで
行くのですか
U：はい
(**知識レベル**:**低 の場合**)
S：円山公園ですと、バス停は祇園になります
(**知識レベル**:**低→説明付加**)
207 系統の東福寺・九条車庫行きのバスは、
3 駅手前の四条西洞院を出発しました。
同様の区間を走る 202 系統の九条車庫・
西大路九条行きのバスは 2 駅手前の
東山三条のバス停を出発しました。
(**知識レベル**:**低→説明付加**)

図 5: ユーザモデル導入後の動作例 2

6 ユーザモデルの判別実験

バス運行情報案内システムの実際の対話データを用いて、ユーザモデル判別の決定木学習を行った。学習には決定木学習アルゴリズム C5.0 を用いた。

6.1 対話データ

本実験で用いた対話データは、2001 年 12 月 10 日から 2002 年 5 月 10 日の間に収集された。電話 (コール) 回数は 215、それに含まれるユーザの合計発話回数は 1492 である。前半はおもに本システム研究関係者が利用し、後半はそれ以外の人も利用している。また、録音された音声データや対話の記録をもとに、主観的に判断された以下のタグが各発話に付与されている。表 2 は、これらの回数の集計結果である。

表 2: 対話データにおけるユーザの分類内訳

	習熟度		知識レベル		性急度	
	発話数	割合 (%)	発話数	割合 (%)	発話数	割合 (%)
低い	743	49.8	275	18.4	421	28.2
不明	253	17.0	808	54.2	932	62.5
高い	496	33.2	409	27.4	139	9.3
合計	1492	100.0	1492	100.0	1492	100.0

- システムに対する習熟度 (高い、低い、不明)
- タスクドメインに関する知識レベル (高い、低い、不明)
- 性急度 (高い、低い、不明)

6.2 実験結果と考察

学習された決定木の判別精度を表 3 に示す。以下は表 3 の 1〜4 の説明である。また、図 6 に 3. で学習された習熟度の決定木の例を示す。

1. 全データを学習データとしたときの closed な条件における判別精度である。
2. C5.0 のオプションである 10-fold cross validation の結果である。全発話をランダムに 10 個のブロックに分け、そのうちの 1 つをテストデータ、残りを学習データとする過程を 10 回繰り返し、テストデータの判別精度の平均をとったものである。
3. 発話ではなくコールをランダムに選び、選ばれたコールに属する発話は全てテストデータとする方が、実情に沿っていると考えられる。そこでコールをランダムに 10 個のブロックに分け、2. と同様の実験を行った。
4. 各ユーザモデルを 3 クラスに分けてタグを付与したが、対話戦略・応答制御に違いが出るものだけ区別すれば十分とも考えられる。そこで、以下のように分類する。
   - 習熟度 { 低い }, { 高い、不明 }
   - 知識レベル { 低い },{ 不明 },{ 高い }
   - 性急度 { 低い、不明 }, { 高い }

   知識レベル以外について、3. と同様の実験を行った。

表 3: ユーザモデルの判別精度

	判別精度 (%)			
	1	2	3	4
習熟度	91.0	80.8	75.3	83.6
知識レベル	91.7	73.9	63.7	-
性急度	90.7	74.9	73.7	90.9

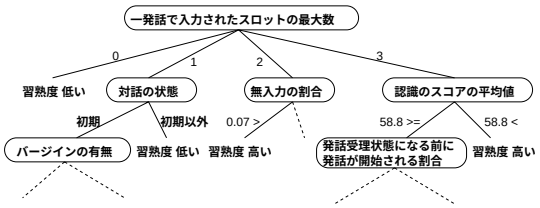

図 6: 習熟度判別の決定木の例

1. の場合が当然、判別の精度が高くなったが、決定木が大きくなってしまい、過学習が起きていると考えられる。

2. と 3. で、判別の精度に大きな差が生じた原因は、2. では個々のユーザに closed な実験になりうること、3. において学習データが不十分であることが考えられる。

3. において十分な精度が得られなかった原因は、学習データが不十分である可能性を除いて、以下が考えられる。まず、判別に利用する特徴が不十分であることが考えられる。判別結果の履歴や、発話速度、ピッチの変化など、今回用いなかった特徴も検討していきたい。次に、特徴を誤って検出してしまうことが挙げられる。発話の誤認識がこの典型例であるが、それ以外にも存在する。例えば、バージインの検出において、ユーザの意図によってなされた場合の他に、雑音による場合がある。また、この場合の判別、学習アルゴリズムとして、決定木が向いているかという問題がある。さらに、対話や判別結果の履歴を用いる方法について再考が必要である。

4. より、今回用いた手法ではシステムに対する習熟度のみが、現時点で効果の期待できるユーザモデルとなった。性急度については、{ 低い、不明 } のデータが全体の 90.7% を占めているため、有意な判別結果ではない。

7 結論

本研究では、ユーザに適応した対話を行うために、音声対話システムへのユーザモデルの導入を行った。京都市バス運行情報案内システムにおいて有効なユーザモデルについて検討し、システムに対する習熟度、タスクドメインに関する知識レベル、性急度の 3 つのユーザモデルを導入した。対話中のユーザの発話内容から得られる特徴と音声の特徴から、ユーザの判別の決定木の学習を行い、その結果に応じた応答が生成されるようにシステムに実装した。

また、実際の本システムの実際の対話データを用いて、決定木学習を行った。その結果、習熟度については 83.6%の精度で判別できた。

参考文献

[1] 日本テレコム Voizi. http://www.voizi.net/.

[2] 大阪ボイスポータル. http://www.vpsite.net/.

[3] D. SADEK. Design considerations on dialogue system: From theory to technology -the case og artimis-. In *Proc. ESCA Workshop on Interactive Dialogue in Multi-modal Systems, Kloster Irsee*, pp.173-187, 1999.

[4] 駒谷和範, 河原達也. 音声認識結果の信頼度を用いた頑健な混合主導対話の実現法. 情報処理学会研究報告, 2000-SLP-30-9, 2000.

[5] 伊藤敏彦, 中川聖一. 音声対話システムにおける協調的応答. 情報処理学会研究報告, 96-SLP-10-19, 1996.

[6] 桐山伸也, 広瀬啓吉. 文献検索をタスクとした音声対話システムの応答生成. 情報処理学会研究報告, 99-SLP-27-16, 1999.

[7] 熊本忠彦. 自然言語対話システムにおける協調的応答の生成. 人工知能学会誌, Vol.14, No.1, pp.3-10, 1997.

[8] 高間康史, 土肥浩, 石塚満. 擬人化エージェントにおける音声対話を通じての協調的応答戦略の自動学習. 人工知能学会誌, Vol.12, No.3, pp.108-116, 1997.

[9] 有田正剛, 島津秀雄. カーナビゲーションシステム用音声対話インタフェース. 人工知能学会研究会資料, SIG-SLUD-9502-1, 1995.

[10] 安達史博, 河原達也, 奥乃博, 岡本隆志, 中嶋宏. Voicexml の動的生成に基づく自然言語音声対話システム. 情報処理学会研究報告, 2002-SLP-40-23, 2002.

[11] W.Eckert, E. Levin, and R.Pieraccini. User modeling for spoken dialogue system evaluation. In *Proc. IEEE Workshop on Automatic Speech Recognition and Understanding*, pp. 80–87, 1997.

[12] 秋葉友良, 田中穂積. ベイジアンネットワークを用いた対話システム: ユーザモデルの推定. 人工知能学会研究会資料, SIG-SLUD-9303-2, 1993.