

# 京都市バス運行情報案内システムの試験評価とユーザモデルの導入

## Experimental Evaluation and User Modeling for Kyoto City Bus Information System

上野 晋一  
Shinichi Ueno

駒谷 和範  
Kazunori Komatani

河原 達也  
Tatsuya Kawahara

奥乃 博  
Hiroshi G. Okuno

### 京都大学 情報学研究科

Graduate School of Informatics, Kyoto University

Our goal is to generate cooperative responses to each user in a spoken dialogue system by choosing an appropriate user model. We analyze real dialogue data collected by Kyoto city bus information system which has been developed at our laboratory. Based on the analysis we implement user models for the system. The categories of user models are skill level to the system, knowledge level of the task domain and the degree of hastiness. The user models are constructed by decision tree learning using real dialogue data collected by the system. Features specific to spoken dialogue systems as well as semantic attributes are used in the classification. We get classification accuracy of about 80% for the three user models. Experimental evaluation shows that the cooperative responses adaptive to individual users serve as good guidance for novice users without increasing the dialogue duration for skilled users. The questionnaire from the users also shows higher performance for the system with the user model.

## 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。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を行うための研究がなされている [Sadek 99] が、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。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をおいて行われてきた [熊本 99]。しかし、音声にはバージョン、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。

本研究では、まず現在我々が開発中であり、試験運用中の京都市バス運行情報案内システム [安達 02] で収集されたデータにより、現在のシステムで行われている対話の評価する。これをもと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、性急度のユーザモデルを導入し、適応的な対話制御・応答生成を目指す。ユーザの判別は、音声や言語、対話レベルなどの特徴を用いた決定木学習により行う。これらを実装したシステムの評価実験により、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

### 2.1 システムの概要

京都市バス運行情報案内システム [安達 02] は、ユーザの指定するバスが乗車する停留所のいくつか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<sup>\*1</sup>。ユーザは乗車場所、降車場所（これらはバス停名か、そのバス停が最寄となる名所・施設）とバスの系統

---

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 系統でよろしいですか  
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

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、図 2 に示す。

現在運用中のシステムの概略は図 3 のユーザモデル導入前のシステムの部分である。この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립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。

#### 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。

### 2.2 システムの試験評価

本システムは現在試験運用中であり、2001 年 12 月 10 日から 2003 年 1 月 22 日まで、全部で 1626 回のコールがあった。このうち、一瞬で電話が切れるようなものを除いた有効なコールは 1244 回であった。これらについて表 1、2 にまとめた。なお、2003 年 2 月には京都市交通局のウェブページに電話番号を掲載してもらったため、それ以降コール数が増加している。

A: 上野 晋一, 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@ar.media.kyoto-u.ac.jp

\*1 公開システムの電話番号: (075) 326-3116

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: 四条河原町からでよろしいですか  
 U: はい  
 S: どの停留所でバスを降りますか  
 U: 嵐山です  
 S: 四条河原町から嵐山までバスでよろしいですか  
 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

表 1: 運用中のシステムでの対話の状況

	コール数	平均対話時間 (秒)	平均ターン数
対話成功	796	*51.5	*5.58
対話失敗	448	37.2	5.28
合計	1244	62.3	7.19

\*: 検索 1 回あたりの平均値

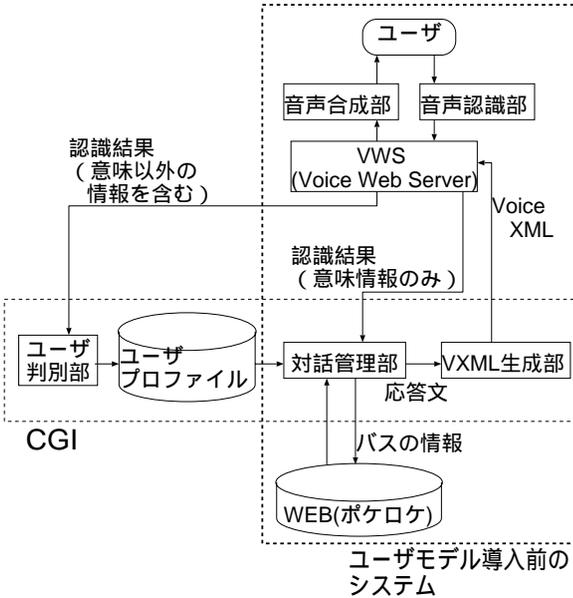


図 3: ユーザモデルを導入したバス案内システムの概略

表 1 では、1 回以上バスの検索が行われたコールを対話成功、それ以外を対話失敗としている。対話成功のコールでは複数回の検索を含むコールが存在するので、1 回の検索あたりの平均を求めた。さらに表 2 に対話時間の分布を示した。

表 1 において、対話失敗では成功に比べ、平均ターン数は同程度であるのに平均対話時間は短い。これはシステムがパーズインを許しているため、フィルラーや雑音が多い場合、ユーザの意図に反してターンが進んでしまい、対話がうまくいかないためであると考えられる。また、表 2 よりユーザによってタスクの達成にかかる時間が大きく異なることが示されており、これらユーザのタイプに応じて応答を変化させる対話戦略が必要となる。

### 3. ユーザモデルを用いた応答の生成

このシステムに対して導入するユーザモデルについて説明する。具体的には、図 3 においてユーザ判別部とユーザプロフィールの部分新たに付け加える。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ィールを対話制御・応答生成に利用する。

表 2: 運用中のシステムでの対話時間の分布

時間 (秒)	~ 30	30~60	60~90	90~120	120~	合計
対話成功	99	475	172	36	14	796
対話失敗	258	131	38	4	17	448

### 3.1 ユーザモデルの分類と対話戦略

導入する 3 種類のユーザモデルそれぞれに関する説明と対話戦略について述べる。

#### システムに対する習熟度

音声対話システムに対するユーザの慣れを表す。判別に有効な特徴は、一回の発話で入力されたスロットの数、パーズインの有無、沈黙の有無などである。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。

#### タスクメインに関する知識レベル

ここでは地域知識の有無についてユーザモデルを用いる。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。認識結果から得られる特徴として、バスの乗車場所・降車場所が正式なバス停名で指定されたかなどを用いる。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。

#### 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。そこで、性急度というユーザモデルを導入する。判別に有効な特徴として、パーズインの有無、沈黙の有無、発話継続時間などが用いられる。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。

### 3.2 決定木によるユーザの判別

ユーザモデルの判別は、決定木を用いてユーザの発話ごとに行う。決定木の学習には実対話データを用いて決定木学習アルゴリズム C5.0 [Quinlan 93] により行う。判別には対話で得られる約 30 の特徴を使用する [上野 02]。これらの特徴には、音声認識の結果による言語情報から得られる特徴に加え、ユーザの発話までの沈黙の時間やパーズインの有無など音声特有の特徴が含まれる。

### 3.3 システムの動作例

ユーザモデル導入後のシステムの動作例を図 4 に示す。最初の発話で入力された項目数が 1 つであることなどから、決定木によりユーザの習熟度が低いと判別される。これにより、対話のフローがシステム主導へと変更される。同様に決定木に

【習熟度：低】【知識レベル：低】【性急度：低】の場合

S: 利用するバス停, 系統番号をどうぞ  
 U: 円山公園  
 S: 順におうかがいします。  
 (【習熟度：低】により対話の流れの説明を付加)  
 乗車場所は円山公園でよろしいですか。  
 (【性急度：低】により対話のフローとして確認を行う)  
 はい, またはいいいえで答えてください。  
 (【習熟度：低】により答え方の説明を付加)  
 U: はい  
 S: 降車場所をバス停名, または  
 最寄りの有名施設を教えてください  
 (【習熟度：低】により答え方の説明を付加)  
 U: 東福寺です  
 S: 降車場所は東福寺でよろしいですか。  
 はい, またはいいいえで答えてください。  
 U: はい  
 S: 検索結果は次のようになります。  
 (【習熟度：低】により対話の流れの説明を付加)  
 円山公園の最寄りのバス停は祇園になります。  
 (【知識レベル：低】によりドメイン情報を付加)  
 202 系統の九条車庫・西大路九条行きのバスは,  
 2 つ手前の東山三条を出発しました。  
 ...

図 4: ユーザモデル導入後の動作例

より性急度が低いことが検知されているため、対話の流れの説明や答え方の説明が付加される。これらは性急度が高い場合にはいずれも省略される。

### 3.4 ユーザモデルの判別実験

2 章で説明したバス運行情報案内システムの対話データの一部を用いて、ユーザモデル判別の決定木学習を行った。

用いた対話データのコール回数は 215、それに含まれるユーザの合計発話回数は 1492 である。これらの各発話に対して「システムに対する習熟度」「ドメインに関する知識レベル」「性急度」の 3 点に関して、「高い」「不明」「低い」のタグを入手により付与した。判別精度を以下の条件で求めた。

- 10-fold cross validation により判別精度を計算する。すなわち、全コールをランダムに 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。
- 習熟度、性急度は発話ごと、知識レベルは対話ごとに判別を行う。知識レベルの判別に用いる特徴が各発話ごとに得られるものではないためである。
- 各ユーザモデルに対して 3 クラスに分けてタグを付与したが、実際の対話戦略・応答制御に違いが出るものだけ区別すれば十分と考えられる。そこで、以下のように分類し、2 クラスへの判別とした。
  - 習熟度 { 低い }, { 不明、高い }
  - 知識レベル { 低い、不明 }, { 高い }
  - 性急度 { 低い、不明 }, { 高い }
- データに付与された各タグの割合の偏りに対処するため、それぞれのクラスのサンプル数の比の逆数を重みとして各データに与えた。C5.0 は判別誤りのデータの数を減少させるアルゴリズムであるが、この重み付けを行うことで、サンプル数の偏りを考慮したうえで誤りの数を減少させる。

表 3: ユーザモデルの判別精度

	習熟度	知識レベル	性急度
判別精度 (%)	85.6	78.2	78.6

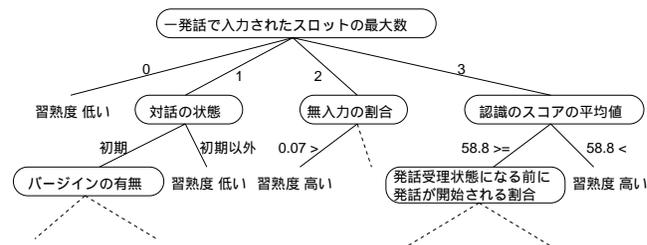


図 5: 習熟度判別の決定木の例

学習された決定木の判別精度を表 3 に示す。また、図 5 に学習された習熟度の決定木の例を示す。

## 4. ユーザモデルを実装したシステムによる評価実験

ユーザモデルを実装したシステムを用いて、被験者による実験を行った。被験者は本システムを使用したことのない本学の学生 20 名である。実験は電話版ではない研究用システムで行い、音声入力に接話マイクを用いた。

### 4.1 実験手順

まず被験者にシステムの概要について説明し、実験条件やシナリオを書いた紙を渡した。8 つのシナリオを 2 セット用意し、ユーザモデルを導入していないシステムと導入後のシステムに対して、シナリオに従いバスの運行情報を検索してもらった。各シナリオは、例えば「あなたは京都に観光に来ています。銀閣寺を訪れたあと、円山公園に行こうと思います。このような状況を想定して実験してください。」のようなもので、実際に入力すべき具体的なバス停名やバスの系統番号などは指定していない。また、性急度を変化させる場合には「携帯電話の通話料を安くすませたいので、できるだけ急いでください。」のような条件をつけた。

アンケート項目には実験前にあらかじめ目を通してもらい、一方のシステムを使い終わるごとに記入してもらった。これはユーザの印象が混同されることを防いだり、ユーザの記憶の負荷を軽減するためである [Over 99]。アンケートは、図 6 に示す 8 項目である。それぞれの項目に対して、例えば「非常にわかりにくかった」から「非常にわかりやすかった」までのような 7 段階のいずれかをチェックしてもらった。

さらに、実験中にシステムから聞いた情報(乗車するバス停名、系統番号、到着までどれくらいかかるか)を紙に記入してもらうようにした。これにより、実験条件を実際の使用状況に近づけている。

### 4.2 実験計画

被験者を 2 つのグループに分け、半数は「ユーザモデル有 → ユーザモデル無」の順にシステムを使用し、半数は逆の順序でシステムを使用した。使用するシナリオの集合は、各グループの半数ずつで前半と後半を入れ替えた。

ユーザの主観的な指標として、各システムに対するアンケート結果を利用する。また対話時間やターン数、音声認識誤りの

1. あなたが話す場面を何を言えばよいかわかりましたか？
2. 対話がうまくいかないときに、わかりやすく誘導してくれましたか？
3. システムの挙動がよくわからず、何が起きているか不安になることはありませんでしたか？
4. 急いでいるときにすばやく情報が得られましたか？
5. バスの案内は余分な情報がなく、分かりやすい説明でしたか？
6. バスの案内は、必要な情報を詳しく説明してくれましたか？
7. システムの使い勝手はどうでしたか？
8. このようなサービスを今後利用したいと思えますか？

図 6: アンケート項目

表 4: 全体の実験結果

		duration	turn
グループ 1 (UM 有 UM 無)	UM 有	51.9	4.03
	UM 無	47.1	4.18
グループ 2 (UM 無 UM 有)	UM 無	85.4	8.23
	UM 有	46.7	4.08

duration: 一回のセッション (シナリオ) の平均対話時間 (秒)  
turn: 一回のセッション (シナリオ) の平均ターン数

回数、ページの回数などの客観的な指標 [Möller 02] もログとして記録しておく。

#### 4.3 実験結果と考察

実験結果をグループごとに分割したものを表 4、5 に示す。グループ 1 は「ユーザモデル有 → ユーザモデル無」の順にシステムを使用し、グループ 2 はその逆の順序で使用した。被験者に対して、どうしてもうまくいかない場合は中断するように指示していたが、タスク達成率は 100% であった。

被験者は本システムを使用した経験がないが、実験の間にシステムに対して急速に習熟していく。表 4 より、あまり習熟していない実験前半では、ユーザモデル有のシステムを使用したグループ 1 の方が対話時間、ターン数が少ない。これは習熟度の低いユーザに対して、提案するユーザモデルにより適切な誘導が実現されており、生成されるシステムの操作方法の説明などにより、ユーザがより早くシステムをうまく使用できるようになったことを示している。一方、システムに慣れたと考えられる実験後半では、2 つのグループの間で対話時間、ターン数に違いが見られない。これはユーザモデル有のシステムにおいて、慣れたユーザに対しては冗長な発話を生成していないことを示している。これらより、システムがユーザに応じた協調的な応答を行っており、習熟度のユーザモデルが有効であることがわかる。

また、表 5 よりユーザの主観的な評価では、ユーザモデル有のシステムが無のシステムより高い評価が得られた。有意水準 5% の  $t$  検定を行った結果、アンケート項目 1、2、7、8 において、グループ 1 の前半とグループ 2 の前半の間で有意差があった。この理由は、ユーザモデル有のシステムで習熟度の低いユーザに適切な応答が行われているためと考えられる。アンケート項目 2、3、4、7、8 において、グループ 2 の前半と後半で有意差があったことから、ユーザモデル無のシステムの応答が習熟度の低いユーザにわかりにくいことが示されている。

#### 4.4 実対話データとの比較

表 2 と表 4 を比べると、表 4 のグループ 2 の UM 無の対話時間から、表 2 の対話成功の 60 秒以上の対話は主にあまり習熟していないユーザが行っていると推測できる。表 4 におけるグループ 2 のユーザに対して、ユーザモデルの導入により対話時間やターン数が大幅に減少したのと同様に、ユーザモデルを導入することで表 2 の長時間の対話の数の減少が期待できる。

表 5: アンケート結果の平均値

グループ 1 (UM 有 UM 無)			グループ 2 (UM 無 UM 有)		
項目	UM 有	UM 無	項目	UM 無	UM 有
1	4.5	4.4	1	3.4	4.3
2	4.5	3.9	2	2.6	3.9
3	3.9	2.5	3	2.6	3.9
4	2.9	2.6	4	2.1	3.6
5	4.2	4.0	5	3.8	4.3
6	3.7	3.8	6	4.5	4.3
7	3.7	3.3	7	2.5	4.0
8	4.5	4.1	8	2.5	3.5

## 5. おわりに

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い、被験者による評価実験を通して有効性を示した。

試験運用中の京都市バス運行情報案内システムの概要を示し、収集された対話データの評価を行った。それをもとにシステムに対する習熟度、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、適応的な応答を実現した。音声特有の特徴を用いた決定木学習により、それぞれのユーザモデルで 8 割程度の判別精度が達成された。さらに、システムにユーザモデルを実装し、20 名の被験者による評価実験を行った。その結果、ユーザモデル実装前のシステムより実装後のシステムにおいて、システムに慣れていない場合に対話時間の減少が見られた。アンケートによる主観的な評価でもユーザモデルを用いたシステムの方が高い評価を得た。

## 参考文献

- [Sadek 99] Sadek, D.: Design Considerations on Dialogue Systems: From Theory to Technology -The Case of Artemis-, *Proc. ESCA workshop on Interactive Dialogue in Multi-Modal Systems* (1999).
- [熊本 99] 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, *人工知能学会論文誌*, Vol. 14, No. 1, pp. 3-10 (1999).
- [安達 02] 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*, SLP-40-23, HI-97-23 (2002).
- [Quinlan 93] J. R. Quinlan, *C4.5: Programs for Machine Learning.*, Morgan Kaufmann, San Mateo, CA (1993), <http://www.rulequest.com/sec5-info.html>.
- [上野 02] 上野晋一, 駒谷和範, 河原達也, 奥乃博: 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, *情報処理学会研究報告*, SLP-42-2 (2002).
- [Over 99] Over, P.: TREC-7 Interactive Track Report, *Proc. of the 7th Text REtrieval Conference (TREC7)* (1999).
- [Möller 02] Möller, S.: A New Taxonomy for the Quality of Telephone Services Based on Spoken Dialogue Systems, *Proc. of 3rd SIGdial Workshop on Discourse and Dialogue* (2002).