

平成 22 年 5 月 1 日現在

研究種目： 若手研究(B)  
 研究期間： 2007 ～ 2009  
 課題番号： 19700164  
 研究課題名(和文) 音声対話システムにおける不確実で曖昧なユーザ発話の理解表現法と応答生成法の研究  
 研究課題名(英文) Study on Ambiguous Understanding Representation and Response Generation for Spoken Dialog Systems  
 研究代表者  
 北岡 教英 (KITAOKA NORIHIDE)  
 名古屋大学・大学院情報科学研究科・准教授  
 研究者番号：10333501

研究成果の概要(和文)：音声対話による情報検索をグラフ探索問題として考え、探索によって誤認識から回復する対話戦略を提案した。グラフは各理解をノードとし、確認なしに誤認識から回復するため、システムは各ターンに複数の理解仮説を保持しつつグラフ全体で最適な理解を“探索”する。適切なシステム応答を選択するために‘情報検索の効率’と‘理解仮説への無矛盾’の基準を導入し、実際のユーザの意図に反する応答による不自然さを排除し、かつ効率的な検索を行う。これらを用いた音声対話システムを構築し、シミュレーション評価と実ユーザ対話評価を行い、ターン数が少なく効率的であること、システムの満足度が良くなる可能性があることを示した。

研究成果の概要(英文)：We proposed a new dialog strategy which can recover from system's misunderstandings without confirmation system utterances. In this strategy, system uses a graph search algorithm, in which each node represent an understanding state. To make a response to the user, system adopts 'efficiency measure' which enables the system efficient search and 'consistency measure' which prevents unnatural response inconsistent with the dialog history. We developed a system with these methods and evaluated it. It is proved that the system was more efficient and obtained more user satisfaction than conventional ones.

## 交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	1,100,000	0	1,100,000
2008年度	900,000	270,000	1,170,000
2009年度	1,000,000	300,000	1,300,000
年度			
年度			
総計	3,000,000	570,000	3,570,000

研究分野：総合領域

科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス

キーワード：音声対話, 誤認識, 複数理解状態, 効率性, 無矛盾性, グラフ探索, ヒューリスティック探索, ユーザ満足度

### 1. 研究開始当初の背景

音声対話インターフェースは人間－機械インタラクションに対して有望であるとされながら、十分に実用化されているとは言い難い。その最大の理由が音声対話の中での誤認識・誤解をどのように扱うかであった。人間も対話のインタラクションの中で相手の発話から自分の誤解を自然に訂正し、対話戦略を変更しつつ対話全体の正しい理解を達成している。人間は誤解しているか否かあいまいな理解状態でも対話を進めていながら徐々に解消するなど、より複雑な対話戦略を用いており、これらを模倣することができなければ自然な対話を実現できないと考えた。

### 2. 研究の目的

対話中のユーザのターン（発話）を認識した結果を複数候補保持し、さらに複数のターンから得られる認識結果のさまざまな組み合わせから複数の発話理解結果を保持しておく、後の誤認識からの回復の可能性を残し、対話を進めるうちに正しい認識・理解結果が選ばれる仕組みを考案すればこれが実現できると考えた。さらに、システムが可能な応答の種類の中から、現在最も「良い」応答として、(1)【無矛盾性尺度】できるだけユーザにとって内容が自然な（ユーザのこれまでの入力内容に矛盾しない）、かつ(2)【効率性尺度】複数の認識・理解を保持することによる曖昧さを効率よく解消する応答を選ぶ戦略が、人間の対話における戦略のある側面を表現していると考え、複数理解の保持と組み合わせで、よりよい対話制御を可能とすることを目的とした。

### 3. 研究の方法

(1) 音声認識の誤認識や発話自体に内在する曖昧性を考慮して、複数の理解候補を常に保持しながら対話を進める方法を考案する。

(2) 上記複数理解の保持方法を、各理解候補をグラフ構造上のノードとみなし、対話の進行をグラフ探索とみなすことで見通しの良い定式化を考案する。

(3) グラフ探索に、最良探索でよく用いられるヒューリスティックスの概念を導入し、それに無矛盾性尺度と効率性尺度を用いることにより、自然な形で応答生成を統合する。

### 4. 研究成果

本研究では、情報検索を、各ノードが理解の仮説候補を表現するグラフ上の探索問題として捉

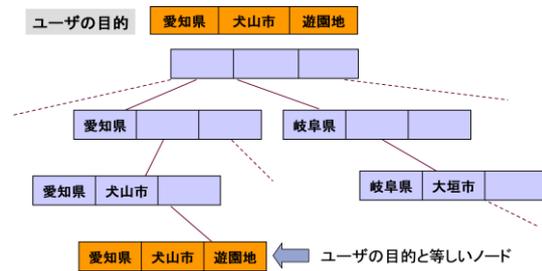


図 1: グラフ探索の例

ランク	都道府県名	市町村名	ジャンル	スコア
1	愛知県	一宮市		1.1
2	三重県	伊勢市		1.0
3	三重県	津市		0.6

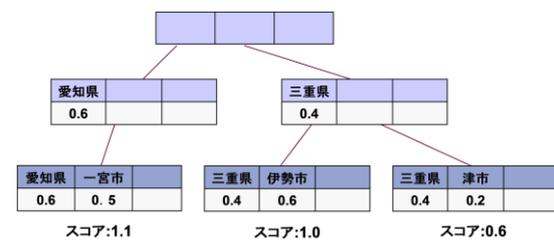


図 2: 複数理解のグラフ表現とベストファースト探索

え、確認発話を減らす新しい対話戦略を提案した。適切なシステム応答を選択するために‘情報検索の効率’と‘理解仮説への無矛盾’に基づいた新しい基準を提案する。それによってシステム応答が実際のユーザの意図と矛盾する不自然さを感じないようにし、また速く検索結果を提示するために検索の効率を考えて応答の生成を行う。さらに、これらの基準をグラフ探索にも利用し、より効率的な対話を実現した。

#### (1) タスク

本研究の目的は検索を目的としたシステム構築にある。そこで、カーナビの目的地検索タスクを採用した。なお、本研究は前半では楽曲検索をタスクとしていたが、手法は同じであり、紙面の都合上ここでは2種類のタスクで同様に動作することでポータビリティがあることを示した旨のみを述べて割愛する。データベースとして昭文社の MAPPLE ガイドデータ（2006年版）を用いた。MAPPLE ガイドデータは「まっぷるマガジン」など発行ガイド情報誌に掲載している魅力ある人気スポット・施設・店舗などの物件情報を収録したデータを含んでいる。物件情報は約 270 種類の分類（ジャンル）から構成される。そこで目的地（物件）の検索のために以下のキーワードを用いること

とする。

- ・都道府県名
- ・市町村
- ・ジャンル（大分類，中分類，小分類）

物件名は 7686 件，都道府県名は 4 件，市町村名は 192 件，大分類は 7 個，中分類は 41 個，小分類は 235 個となっている。またこのキーワード間には階層構造が存在し，言語理解にも用いられる。システムはそのキーワードに対して質問，確認，検索結果の提案を行うものとする。

## (2) グラフ探索に基づく情報検索のための音声対話理解

### ① グラフ探索としての音声対話理解

まず始めに，情報検索の過程をグラフ探索であると考え，ここで音声対話などを通じて検索キーをスロットに埋めていく過程をグラフで表現することを考える（図 1）。ユーザは検索を行うために目的（あいまいかもしれないキーワード集合）を持っている。グラフのノードとして（部分的な）キーワード集合を考えることで，理解の探索グラフを構成することができる。探索の過程において，アクティブノードは現在の理解状態を表し，システムは各対話ステップから音声認識結果に基づいてノードを展開していく。誤った探索が進む場合もありうるが，何らかの探索手法においてバックトラッキングを行って回復することも可能となりうる。

### ② 複数理解の適用

本研究の研究代表者らの以前の研究において，システムは N-best の認識仮説を用いて複数の理解仮説を保持し，正解の理解仮説が優先されるように適切なシステム応答を選択する対話戦略を提案した。この論文の手法において，図 2 の上に示すような，複数のアクティブノードの存在を許すことは，下のように複数の理解仮説を保持しておくことと等しい。このとき新たな入力があったとして，その N-best の音声認識結果を使うと，各アクティブノードは最大で N 個の新しい後続のノードに展開される。

### ③ ベストファースト探索

探索にはベストファースト探索を用いる。すなわち，ノードスコアリング戦略を使って展開するノードを決定する。例えば，われわれのシステムにおいては，認識結果から得られる単語の信頼度スコアを使用する。各ノードはノードに含まれる単語の信頼度スコアの和をスコアとして持っている。このスコアを探索過程における探索スコア  $g(n)$  と呼ぶこととする。図 3 のグラフでは，左端のノードを選択し，そのノードの内容に基づいて新しい情報を聞き出す質問を生成し，そしてユーザの応答の認識結果を受けてノードの展開を行う。 $g(n)$  を使用するだけでなく，ノードを評価するために「発見的な」（ヒューリスティック）知識を導入することもでき

る。これをここではヒューリスティックスコア，この手法は探索スコア  $g(n)$  とヒューリスティックスコア  $h(n)$  の両方を用いて最も高いスコア  $g(n) + h(n)$  を持つノードを選択する。探索問題としては，典型的な  $h(n)$  は未来の  $g(n)$  の予測であるが，ここでは本当の意味でのヒューリスティックなスコアを使用する。このスコアは次の章で提案するシステム応答（質問）の生成の基準として用いるスコアを利用する。

## (3) システム応答選択のための基準

本研究では，理解仮説に対する無矛盾性と，情報検索における効率の組み合わせに基づく新しい基準を提案した。確認発話は確認に対する答えが，多くの場合（つまり，第一ベスト仮説が正解である場合）においてベストな仮説以外全ての仮説を棄却できる条件のもとでは最も良い応答かもしれない。しかし第一仮説に依存する発話は本当の状態と矛盾しているかもしれない，ユーザが応答を不自然に感じたりシステムの誤認識に気づくという状況になるかもしれない。ユーザに誤認識を気づかれることなく誤認識から回復するために，無矛盾の観点からできるかぎり多くの理解仮説と矛盾しない発話が好ましい。加えて，情報検索を行う上でシステムはできる限り早くユーザに目的地を提案することが好ましい。もしシステムが 2 つの質問の選択肢を持っているなら，検索結果を絞り込める質問が望ましい。これら 2 つの観点を考えて，システムはその時点での適切なシステム応答を選択する必要がある。

### ① 理解仮説との無矛盾尺度

図 2 で表現された理解仮説による理解状態の場合を考える。もしシステムが“愛知県でよろしいですか?”と質問した場合，この質問は 2 番目と 3 番目の理解仮説と矛盾する。これはユーザが三重県と発言したにもかかわらず，誤認識によって愛知県が得られた場合があるからである。また同様に“一宮市でよろしいですか?”という質問も 2 番目と 3 番目に対して矛盾した質問となる。このような過去のユーザ発話との矛盾の可能性が高く，誤認識がユーザに明らかになる可能性の高い発話を避けるために，つぎの無矛盾の尺度を適用する。

$$S_c(q) = 1 - I(q, n)$$

ここで質問  $q$  が理解仮説  $n$  と矛盾する場合， $I(q, n) = 1$ ，それ以外の場合  $I(q, n) = 0$  である。

### ② 情報検索における効率尺度

情報検索タスクにおいて，探索空間を大きく絞り込める質問は効率的である。まったく探索空間を絞り込めない質問というのは単に時間の無駄となってしまう。そこで質問を行うことによってどれだけのエントロピーを減少できるかを

推定するために検索効率尺度として相互情報量を使用した。複数の理解仮説に対する相互情報量は以下のように定義される:

$$S_s(q) = I(X; q | n) = H(X | n) - H(X | q, n)$$

ここで、 $H(X|n)$  と  $H(X|q; n)$  は以下のように定義されている:

$$H(X | n) = - \sum_{x \in X} p(n, x) \log_2 p(x | n)$$

$$H(X | q, n) = - \sum_{a \in A_q} \sum_{x \in X} p(n, x, a) \log_2 p(x | n, a)$$

ここで  $A_q$  は質問  $q$  をした後、ユーザから得られる可能性のある回答の集合であり、 $X$  はデータベースからの検索結果の集合である。

### ③ 最終的なシステム応答決定

最後に、上記の2つの尺度のバランスを決定しなければならない。ここでは、システムは次の式のように重み付け和の最大値を持つ質問  $\hat{q}$  を選択することとした。

$$\hat{q} = \arg \max_q \{w_c S_c(q) + w_e S_e(q)\}$$

### ④ ヒューリスティックスとしての尺度の使用

これらの尺度をヒューリスティックス  $\hat{h}(n)$  として利用した。ヒューリスティックスを用いることで、探索スコア  $g(n)$  だけを使用するより目的に早く到達しようという効率を考えて探索を行うことができると考えられる。ヒューリスティックスを用いてベストファースト探索を行う場合は、以下の式を評価して探索を行うノード  $\hat{n}$  を決定する。

$$\hat{n} = \arg \max_n \{g(n) + \hat{h}(n)\}$$

ここで、今回はヒューリスティックスの計算方法として、ローカルに一つの仮説だけを考える場合とグローバルに理解仮説全体を考えて計算を行う2通りの方法を考える。

ノード  $n$  だけを考慮したローカルなヒューリスティックスは、上記のように求められる質問  $\hat{q}$  によって以下のように求められる。

$$\hat{h} = w_c S_c(\hat{q}) + w_e S_e(\hat{q})$$

一方、グローバルに理解仮説全体を考えて計算を行うこともできる。これは、最大スコアをもつ  $n$  のみではなくアクティブノードの集合  $N$  も考慮して計算が行われるため、この章で提案したスコアを修正する必要がある。まず無矛盾尺度を、次の式のように全体に対して計算を行う。

$$S_c(q) = \sum_{n \in N} (1 - I(q, n)) P(n)$$

ここで、 $P(n)$  は理解仮説  $n$  が正しい確率であり、このスコアは正しいと思われる理解仮説を優先するための重みとして使用される ( $P(n)$  は本来事前に推定しておくべきだが今回は簡易法(省略)を用いた)。また、検索効率尺度も、以下のように全体を考慮して計算する。

$$S_e(q) = I(X; q | N) = H(X | N) - H(X | q, N)$$

ここで、

$$H(X | N) = - \sum_{n \in N} \sum_{x \in X} p(n, x) \log_2 p(x | n)$$

$$H(X | q, N) = - \sum_{n \in N} \sum_{a \in A_q} \sum_{x \in X} p(n, x, a) \log_2 p(x | n, a)$$

である。グローバルなヒューリスティックスを用いた場合には、全体のアクティブノードを考慮するので、対話の過程においてバックトラッキングが生じた場合に急な話題の飛躍が生じる可能性を減らせる可能性がある。

### (4) 評価実験

#### ① シミュレーションによる客観評価実験

提案する対話管理を対話の平均ターン数によって評価を行った。本稿では、自動的にユーザ発話を生成することでコンピュータによるシミュレーション対話によって評価を行った。3つの提案戦略(ヒューリスティックスを用いないものと2種類のヒューリスティックスを用いるもの)と典型的な手法(ターンごとに毎回確認を行う手法、尤度に基づいた確認戦略)とを比較した。シミュレーションは以下の手順で行った。

1. 始めにユーザシミュレータがランダムに目的(キーワード集合)を決定する。以下、対話によってユーザシミュレータは次の手順を通してこの目的の設定を完了しようと試みる。
2. システムが最初の発話、“目的地を検索します。条件をおっしゃってください。”を行う。
3. ユーザシミュレータがシステム発話に応じて返答を行う。
4. ユーザシミュレータの発話から事前に決められた認識率を基に、擬似音声認識器が擬似的な認識結果とその信頼度スコアを作成する。
5. システムはその擬似的な認識結果を使ってアクティブノードを展開し理解仮説を更新する。
6. もし理解状態が終了条件を満たしていない場合、システムは次の発話を生成し、ステップ3に戻る。

ステップ4において、音声認識率  $R$  を60-100%と事前に定義した。提案システムにおいて、全てのスロットが埋まった場合、検索条件にあった目的地の提案を行う。もし目的地が事前に設定されたユーザの目的とマッチする場合、ユーザはシステムの提案に対して肯定発話をし、対話終了とする。それまで、システムは質問と提案をし続けるものとする。上記のシミュレーシ

手順のステップ6で、(2)と(3)で説明した手法を提案手法として用いた。ターンごとに毎回確認を行う戦略において、システムはユーザの新しく入力された情報に対して確認発言を行うか、またはユーザからの確認に対して“はい”という肯定発言を受けた後、新しい情報に対す

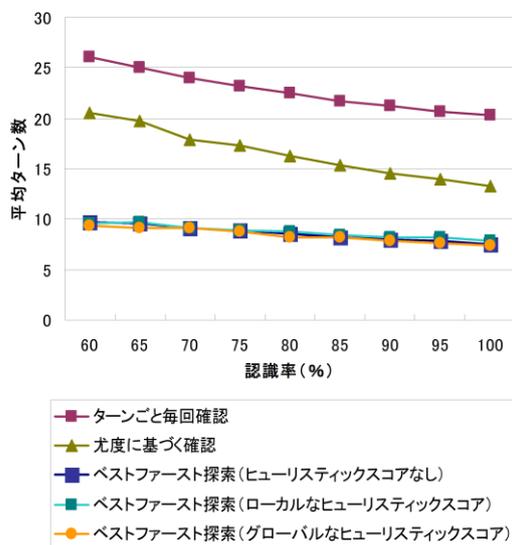


図 3: シミュレーションによるターン数評価

る質問を行った。すべての手法において、ユーザシミュレータは誤認識に気づいたとき訂正発言によって誤認識からの回復を行った。各手法に対して1000回のシミュレーションを行った。実験結果を図3に示す。この結果から、提案戦略を用いることでどの認識率においても平均ターン数を減少することができた。若干ではあるがヒューリスティックスコアを用いることでターン数はより減少した。

## ② 実ユーザによる主観評価実験

次に実ユーザに実際に音声対話システムを使用してもらい対話戦略の評価を行った。比較する対話戦略は従来手法として、毎回確認を行うものと尤度に基づいて確認を行うもの、提案手法としてグローバルなヒューリスティックスコアを用いたベストファースト探索の3手法とした。各戦略ごとに6対話ずつ合計18対話を、6名のユーザに評価を行ってもらった。評価基準は、以下の4項目に関して1(悪い) - 5(良い)の5段階評価してもらった。

- ・システムの満足度
- ・システムの賢さ
- ・対話の自然さ
- ・目的達成までの効率の良さ

実験ではあらかじめ与えられた検索のためのキーワード対を基に対話をしてもらい、その条件にあった検索を完了するまで対話を行ってもらった。シミュレーションでは1発話1スロットの

発言としていたが、この実験ではユーザは1発話複数のスロットにわたる発言を許可した。結果を図4に示す。提案手法では、N-bestの認識結果を用いているため、正しい認識結果が第1候補にあるにもかかわらず、挿入誤りのある候補に高いスコアを与えてしまう場面があった。その

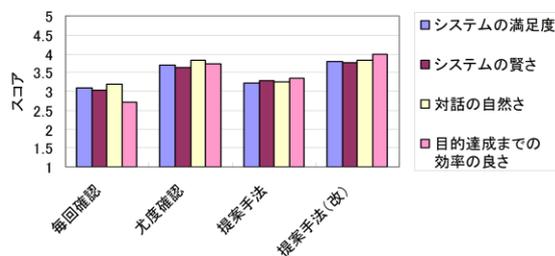


図 4: ユーザの主観評価結果

結果、システムは誤った提案をしてしまう場面が見られた。また、ユーザは、システムの誤った提案の対処方法が分からず、誤認識から抜け出せない場合があった。これらの状況は、ユーザの満足度を著しく低下させてしまう結果となった。そこで、これら例外的な状況を除いた結果を、提案手法(改)として提示する。この結果は、ある特定の状況になったときのシステムのふるまいを工夫することで実現可能となる。このとき提案手法は他の手法より、平均ターン数の減少に加え、良い主観評価結果を得ることができることが明らかとなった。

本成果は音声関係でも権威ある国際会議 ICASSP にも採録され、多くの世界的研究者にも知られるところとなった。また、本研究は現在世界中で精力的に研究がなされている POMDP

(Partially Observable Markov Decision Process) による方法とも関連が深く、かつ容易に実装可能である。今後より数学的精緻化をしていくことによりスタンダードな方法となることを目指す。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[学会発表] (計 4 件)

- ① Y. Kinoshita, C. Miyajima, N. Kitaoka, K. Takeda, "Spoken dialog strategy based on understanding graph search," ICASSP2009, 4797-4800, Apr. 2009. (Taipei International Convention Center, Taiwan)
- ② 木下 裕司, 宮島 千代美, 北岡 教英, 武田 一哉, "理解状態グラフの探索による音声対話戦略," 音響学会春季研究発表会 3-5-3, 4 pages, Mar. 2009. (東京工業大学)
- ③ 木下 裕司, 宮島 千代美, 北岡 教英, 武田 一哉. "理解状態のグラフ探索に基づ

いた音声対話戦略,” 電子情報通信学会  
研究報告 NLC2008-19, pp. 1-6, Nov. 2008.  
(NiCT(東京))

- ④ 杉本 夏樹, 矢野 浩利, 北岡 教英, 中川  
聖一. “誤認識に対処した自然で効率的  
な音声対話戦略の評価,” 情報処理学会  
研究報告, 2008-SLP-70, pp. 181-186, Feb,  
2008. (伊東, 静岡)

## 6. 研究組織

### (1) 研究代表者

北岡 教英 (KITAOKA NORIHIDE)

名古屋大学・大学院情報科学研究科・

准教授

研究者番号 : 10333501