

地図情報検索のためのマルチモーダルインタフェース

- 検索対話制御方式とそのプランによる実現の検討 -

瀬川 修[†], 田村 直良[‡], 中川 裕康[‡]

[†]中部電力株式会社 電力技術研究所, [‡]横浜国立大学 工学部 電子情報工学科

1 はじめに

本研究では、情報検索のための知的な“仲介者”を対話システムとして実現することを目的にしている。近年、人工知能の分野では計算機システムを知的主体とみなすことで、対話によるインタラクションの研究が盛んに行われている。ユーザ/システムの言語指向による対話手段としては、コマンドによるオペレーションが一般的であるが、計算機アーキテクチャに依存しない人間にとって普遍的な自然言語による対話の試みもなされている。自然言語入力によるインタフェースは、計算機OSやアプリケーションに依存したコマンド体系、操作手順に習熟する必要がなく、思考のプロセスを直接コマンドとして反映できるなどのメリットがあると言われている。しかし、自然言語入力がコマンドに対して不利な点として、曖昧性、冗長性に起因して簡潔な論理表現が困難であること、機械表現とのミスマッチなどがあり、計算機との対話手段として用いるためには“仲介者”がこれらの意味的ギャップを解消しなければならない。すなわち、自然言語入力は“仲介者”が十分に知的であるという前提のもとに初めて成り立つ（実用に耐えうる）ものである。

一方、言語指向に加えてポインティングデバイス等による直接操作をうまく取り込んだマルチモーダルインタフェースに関する研究が盛んに行われている[1][2]。これは言語入力とポインティングデバイスなどの非言語入力の併用によって、単一モードのウィークポイントを相互に補完するというコンセプトであり、自然言語インタフェースの実装としても有効な方法の一つであるといえる。

本稿では、2次元視覚情報（街路図面）を用いた情

報検索タスクにおいて、対話の構造を漸次的に理解する枠組の検討を行なう。ユーザ/システムの目的指向の対話の構造を捉え、マルチモーダル・インタラクションにおける入力の漸次的な理解と対話制御方式について詳細な検討を行ない、そのプランによる実現の一手法を提案する。

2 検索インタフェースの概要

2.1 検索タスクについて

データ検索のアプリケーション例として電力会社で用いられる地図情報システム、すなわち、電子化された街路図面をベースとしたビジュアルな入出力が可能な検索システムを対象としている。設備データ（電柱、変圧器、管路 etc.）は地図座標上のオブジェクトとして表示される。検索タスクの特徴として、視覚情報を用いることから言語入力とポインティングデバイス¹を用いた直接操作による検索条件指定が併用できることが挙げられる。図1に地図情報システムにマルチモーダル検索インタフェースを適用したイメージを示す。対象ユーザ（電力会社社員）としては以下の設定を行っている。

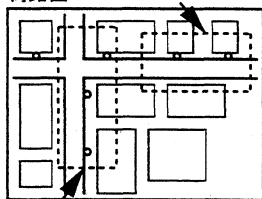
- 設備名や検索条件を（断片的でもよいから）自然言語で表現できる。
- データ構造、DBMSや検索言語に関する知識は前提としない。

2.2 知的システムにおける検索スタイル

図1に示すように、我々はユーザから見て“知的”なインタフェースを介した情報検索のスタイルは、

¹ 言語入力的手段としてキーボード、ポインティングデバイスとしてマウスを用いる。

「ここにある1970年以前
の街路図の変圧器を表示しろ。」



「この範囲ではどうか。
同じ条件で探せ。」

検索対話例

ユーザ : D1 ここにある1970年以前の変圧器を表示しろ。
システム: D2 〇〇件あります。 D3 出力条件を入力して下さい。
ユーザ : D4 色替え表示。
システム: D5 <結果を画面に表示>
ユーザ : D6 この範囲ではどうか。同じ条件で探せ。
システム: D7 該当する設備はありません。
ユーザ : D8 1971年以前では。
システム: D9 〇〇件あります。<結果を画面に表示>

「」 言語入力

➡ ポインティングデバイス入力 (ポイント、領域指定)

図 1: インタフェースの適用イメージ

従来のシステム主導のメニュー型インタフェースによるものと比較して大きく異なると考えている。システムの実現にあたっては、ユーザ主導による柔軟な「思いつくまま」の検索など、自然言語入力のメリットを最大限に享受できるインタフェースを提供しなければならない。

しかしながら、自然言語入力は、その曖昧性に起因した論理表現の困難さから、システムに対し検索要求を1文、ないしは1度の検索ターンで条件として過不足なく与えるのは大変困難である。また検索条件によっては、言葉で表現するより直接操作のほうが効率がよい場合もある。従って、ユーザ主導による知的インタフェースを指向する場合には、検索要求はユーザの1度の入力で決定されてしまうのではなく、比較的単純な(断片的な)条件の積み重ねをシステムとのマルチモーダル・インタラクションにより進めていき、タスクを達成するといった検索スタイルへの転換が求められる。

3 検索対話システムの枠組

目的指向対話の一つである検索対話において、どのようにしたら協調的な対話の構造²が生まれるのであろうか。ユーザ/システムから成る系は、対話の開始から終了までのイベント列の入力によって刻々と変化し続ける「状態遷移機械」に他ならない。協調的な対話とは、いわゆる「対話のルール」である状態遷移のシーケンスに従いインタラクションが行な

われることである。対話システムの設計とは、本質的にこのような「状態遷移機械」を作り出すことであるとも言える。

以下では本研究で目的とする検索対話システムを実現するための重要なコンセプトを説明し、マルチモーダル・インタラクションを対象とした対話制御の実現方式を明らかにする。

3.1 入力の取り扱い—イベント駆動

ユーザの検索要求は、2つのモードから入力される断片的発話とジェスチャから成るイベント列である。言語モードにおいては、発話はいわゆる書き言葉のような完結した1文単位とは限らないため、入力をイベント単位として捉える必要がある。すなわち、2つのモードの入力イベントの時系列は同等の扱いが必要であり、システムは常にイベント列の発生を監視しながら、発話ターンの中でイベント列の意味解釈を随時行なわなければならない。

言語入力的手段としてキーボードを使用する場合は、入力デバイスの特性から2つのモードは必然的に交互入力となる。交互入力では、発話とジェスチャのイベント発生の時間的な同期の情報を明示的に得ることは一般に困難あり、発話ターンの区切りの認識は容易でない。本稿では、本来の意味でのイベント駆動とは異なるが、簡単のためユーザが発話ターンの開始と終了を明示的にシステムに通知を行なうものとする。

² ユーザがシステムの振舞いを協調的、自然であると感じるような心的状態のこと。

3.2 対話構造の漸次的な理解

マルチモーダル・インタラクションにおけるイベント列の統合的な意味解釈を行なうためには、個々の発話やジェスチャの独立した解析だけでは不十分であり、状況に依存した文脈情報、すなわちマクロな対話構造の理解が必要である。マクロな対話構造は、対話セッション終了後に事後的に見えてくる(確定する)ものである。すなわち、リアルタイム処理を指向するシステムにおいては、対話構造の解析を対話セッション終了後まで待ってから事後的に行なうことができないため、対話進行中に「意味構造」の漸次的な理解と生成が必要とされる。さらに、目的指向の対話を協調的に進めるためには、システムは、ある発話ターンにおいてユーザの状態とシステム自身の状態(内部状態と呼ぶ)を各々把握しておく必要がある。内部状態は発話の主導権の管理など、後述する協調的な対話制御のプランニングで使用される。

全体の対話制御の仕組みとしては、対話の進展に伴いボトムアップ的に構造を生成しながら、発話ターンが移った時点で、ユーザとシステム自身が各々の状態にいるかを把握し、その状況下での最適なプランニングを行なうという方策を用いる。

3.2.1 内部状態の遷移

検索対話のセッションが続く間、ユーザ/システムからなる系は入力イベントに同期して内部状態の遷移を繰り返す。図1に示す検索対話例(D1~D9)について、内部状態の遷移の例を表1示す。

表1: 内部状態遷移の例

user	system	interaction
(start)	wait	-
talk	listen	→ D1(情報要求)
wait	planning	-
listen	talk	→ D2(情報提示), D3(情報要求)
planning	wait	-
talk	listen	→ D4(情報提示)
wait	planning	-
listen	talk	→ D5(情報提示)
planning	wait	-
talk	listen	→ D6(情報要求)
wait	planning	-
listen	talk	→ D7(情報提示)
planning	wait	-
talk	listen	→ D8(情報要求)
wait	planning	-
listen	talk	→ D9(情報提示)
(end)	wait	-

3.2.2 検索対話の構造

検索対話は、基本的にユーザとシステムの発話の主導権の移動をともしなうインタラクション(発話ターン)の連続である。図2にタスクの分析の結果定義した検索対話の構造を示す。最上位レベルの「検索セッション」は、ある一つの検索タスクを達成するための対話の単位である。

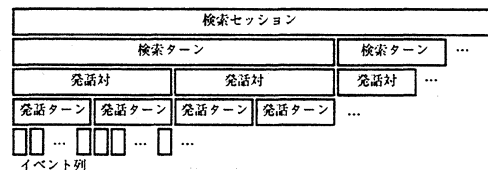


図2: 検索対話の構造

4 プランによる実現

図3にシステムのアーキテクチャの構成を示す。ここでは、前節で述べた内容を具体化し、対話制御モジュールのプランニングについて説明する。

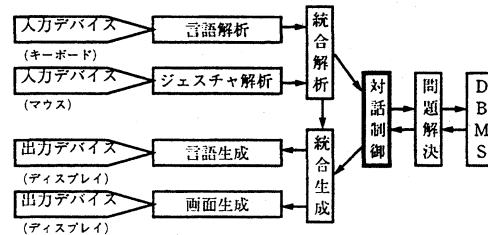


図3: システム構成

4.1 プランの記述

4.1.1 スロットの定義

Condition はプランの適用条件を、**Effect** はプラン実行後の効果を、**Action** はプラン適用時の実行内容を記述する。スロットの定義(スタックはユーザの要求タスクの管理を行なう)を表2に示す。

表2: スロットの定義

Condition	カレントの対話構造 ユーザの内部状態 システムの内部状態 スタックの状態
Effect	内部状態の遷移
Action	他のモジュールの起動 対話構造の更新 内部状態の更新

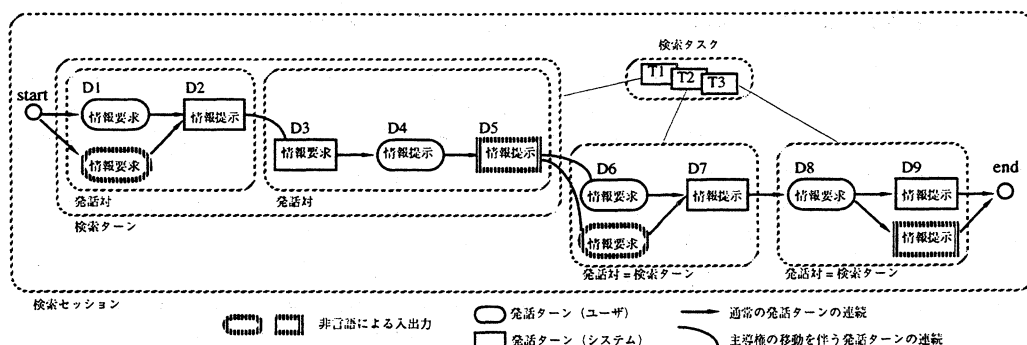


図 5: 生成された対話構造の例

表 3: プランの記述例

Description	fulfill user's request
Condition	対話構造の状態: 接続可、主導権ユーザー ユーザーの内部状態: wait システムの内部状態: planning スタックの状態: push
Effect	内部状態の遷移: user(wait → listen) system(planning → talk)
Action	問題解決モジュールの起動

4.2 動作の一例

対話構造は検索セッションの開始(start)から、検索タスクが充足し対話の終了(end)に至るまで漸次的に生成されていく。図1に示す検索対話例について、動作例を図4に示す(D6のターンについて例を示す)。この場合、表3に示すプランが適用される。

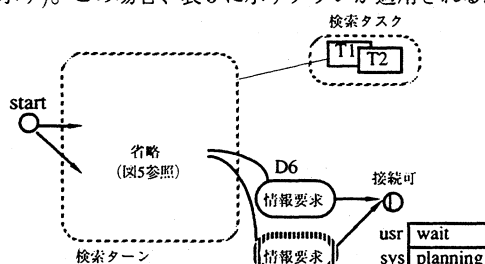


図 4: プランニングの動作例

このようなプランニングによって、最終的に生成された対話構造の例を図5に示す。

5 おわりに

本稿では、情報検索のための知的インタフェースの実現を目的として、検索対話の構造を漸次的に理

解する枠組の検討を行なった。地図情報検索におけるユーザ/システムの目的指向対話の構造を捉え、マルチモーダル・インタラクションにおける入力イベントの漸次的な理解と対話制御方式について検討を行ない、そのプランによる実現の一手法を提案した。今後の課題としては、対話システムへのプランニングの実装によってモデルの妥当性の検証を行なっていく予定である。

「使える自然言語インタフェース」を構築するためには、言語解析処理などの頑健性向上も重要であるが、(1) タイムアウトの概念導入、(2) 要求の割り込み/インタラプトの処理についても検討しプランニングの機能向上を図っていく必要があると考えている。また、実世界におけるより柔軟なインタラクションを扱うために、熟考と即応による動的なプランニングの枠組 [3][4] の検討も同時に進めていく必要がある。

参考文献

1. Kyung-ho Loken-Kim, Suguru Mizunashi, Mut-suko Tomokiyo: "ANALYSIS AND INTEGRATION OF MULTIMODAL INPUTS IN INTERPRETING TELECOMMUNICATIONS", 情処研報 95-SLP-7 Vol.95 No.73 pp.73-78 1995-7
2. 島津秀雄, 高津洋典: "Multi-Modal Method - マルチモーダルインタフェース構築方法論 -", 情処研報 95-SLP-5 Vol.95 No.16 pp.9-16 1995-2
3. 瀬川修, 田村直良: "エージェントモデルによるマルチモーダル検索インタフェースの提案", 言語処理学会第2回年次大会論文集, pp.117-120 1997-3
4. 長谷川隆明, 中野有紀子, 加藤恒昭: "反射と熟考の相互作用に基づく協調的対話モデル", 信学技報 NLC96-27 Vol.96 No.158 pp.43-50 1996-7