

認識モデルとDPマッチングを用いたマルチモーダル入力からの戸惑い検出

3 B - 4

龍田 幸拓・田野 俊一

電気通信大学大学院 情報システム学研究所

1 はじめに

人間とコンピューターのコミュニケーションにおいて、様々なモダリティーを統合して使用することで、より自然なコミュニケーションを可能にするマルチモーダルユーザーインターフェース (MMUI) に関する研究が盛んに行われている。

また、コンピュータの性能が向上することによって、機械の高機能化が進み、様々な処理が可能となってきたが、逆に、操作が複雑になり、ユーザーにとっては、決して使いやすい機械ばかりではないのが現状である。

本研究では、様々な情報をマルチモーダル (MM) 入力から取得し、ユーザーの意図を推定することで戸惑いを検出し、ユーザーをサポートするシステムの構築を目指す。

2 研究課題

MMUIに関する研究では、音声とタッチパネルを用いた航空券予約システム[1]や音声とジェスチャーを用いたマルチモーダル対話システム[2]、形式文法による解析アルゴリズム[3]など、多数の研究がなされており、複数の入出力モダリティーを使用することでより自然な入力が可能となったり、認識率が上がるなどの効果が見られる。

これら従来の研究では、単にユーザーの動作などのマルチモーダル情報を統合認識し、推定されるコマンドを実行するシステムが多い。

しかし、MM入力を利用すれば、ユーザーの思考状態など、様々な情報を得ることができる。

本研究では、従来のMMシステムで出力される実行命令に加えて、マルチモーダル情報をもとにユーザーの思考状態を推定し、ユーザーの戸惑い[4]を検出する。

3 アプローチ

3.1 対象

本研究では、ATM（銀行などの自動支払機）を対象とする。ユーザーの視線や手の動きなどのMM入力情報を取り込むことによって、例えば、『「カードを挿入して下さい」と機械から指示されたが、挿入され

ず、視線がさまよっている場合に、カード挿入位置を教える』といったサポートを可能とすることが目的である。

3.2 基本方式

本研究では、図1のように人間と機械をそれぞれモデル化した。これを推定モデルとして利用することにより、典型的パターンを導出する。次に、導出した典型パターンと観測されたユーザーのMM入力の間でDPマッチングを行い、ユーザーの動作・思考を推定しサポートするというアーキテクチャを設計した。

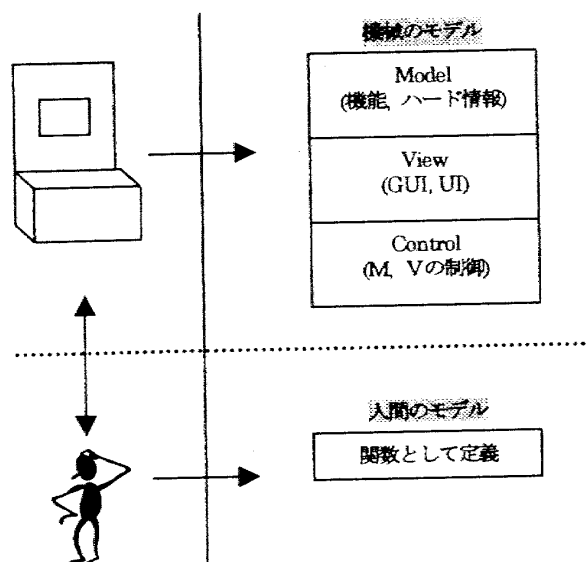


図1 推定モデル（機械と人間のモデル化）

3.3 モデル表現(図1参照)

(1) 機械のモデル

操作対象の機械を Model・View・Control (MVC) に分けてモデル化した。Mには機械の機能やハードウェア情報を記述し、Vには画面の情報やインターフェース部分に関する情報を記述する。CはMとVの間での制御情報を記述する。これらは図2に示すような文法により、記述する。そして、MVCを結合することによって、図3に示すようなベトリネットを基にしたモデルで対象とする機械が表現できる。

(2) 人間のモデル

人間の行動に関するモデルを関数として記述する。例えば、見る場所の移動による各部（視線・頭・手・

体)の移動・確認による各部の移動・機器操作による各部の移動といった形で定義する。実際には、人間工学や、人間の行動・動作などに関する研究を参考にする必要はあるが、現在はアドホックに定めている。

```

<モデルファイル> ::=
  (<モデルノードの定義> | <トランジションの定義>
   | <アークの定義>)...
<モデルノードの定義> ::=
  node<CR>
  name <名前><CR>
  input <入力名> (<入力名>)...<CR>
  output <出力名> (<出力名>)...<CR>
  (mprocess <関数名><CR>)|<CR>
  retvalue <戻り値>, <出力名>
  (<戻り値>, <出力名>)...<CR>
  end<CR>

```

図2 文法記述例

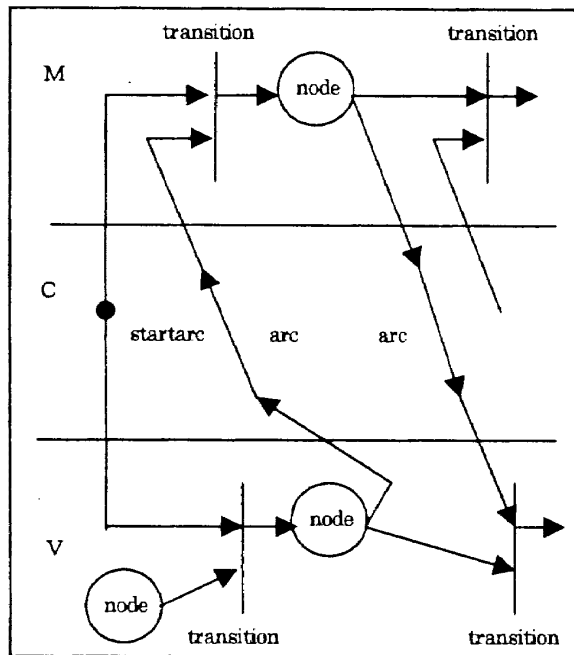


図3 モデル例

3.4 DPマッチングと戸惑い検出

機械のモデルと人間のモデルを分析することにより、対象とする機械を利用する際の人間の典型的行動パターンを生成できる。このパターンはマルチモーダル情報を持っている。モデルを用いて導出しているので、各典型パターンに対し、各々の時刻において、どのモダリティーがどのくらい重要かに関する情報も導出可能である。そこで、このモダリティー間の重要度に関する指標を有する典型パターンと、ユーザーの実際の操作パターンを、DPを用いてマッチングさせ、ユーザーの戸惑い状態を推定する。

マッチングの結果は図4のように分類することができる。横軸は入力パターン列、縦軸が典型パターン列である。結果が、D3からD6の場合は、何らかの異

常がある場合であり、ユーザーが戸惑っているか、もしくはユーザーの操作が誤っているということになる。D3からD6の場合には、マッチングの結果から原因を絞り込むことによって、ユーザーをサポートすることができる。

- ・ D1, D2: 正常な場合。D2は入力が冗長。
- ・ D3: 途中までマッチしているが、途中からあるモダリティーがマッチしなくなった場合。
- ・ D4: 途中までマッチしているが途中から全然マッチしなくなった場合。
- ・ D5: 最初はマッチしないが、途中からマッチしている場合。
- ・ D6: 最初はマッチして、途中でマッチしなくなり、最後にまたマッチした場合。必要な操作が抜けている。

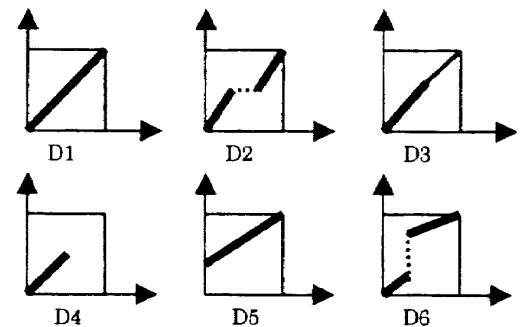


図4 DPマッチングの結果

4 おわりに

現在、MM入力として、視線・タッチパネル・ペンパッドを使用し、認識モデルを用いて戸惑いを検出するプロトタイプシステムを構築し評価を進めている。

今後は、リアルタイム化のため、DPマッチングの改良による計算時間の短縮が必要である。また、音声やジェスチャなど、より多くのモダリティーに拡張することによって、もっと正確にユーザーの行動、思考を推定することができると思われる。

参考文献

- [1] Laurence Nigay, Joelle Coutaz, "A Generic Platform for Addressing the Multimodal Challenge", HUMAN FACTORS IN COMPUTING SYSTEMS, CHI'95 CONFERENCE PROCEEDINGS, ACM PRESS, pp. 98-105, (1995).
- [2] 湯浅 夏樹, 三谷 純司, 外川 文雄, "マルチモーダル対話データベースに基づく音声と身振りの認識統合モデル", 情報処理学会論文誌, Vol. 37, No. 6, pp. 1049-1060, (1996).
- [3] 若梅 友則, 田野 俊一, 形式文法によるマルチモーダル入力の表現と解析, 第11回人工知能学会全国大会, pp. 392-395, (1997).
- [4] 紙谷 卓之, 丸谷 洋二, "ワープロ作業における困り状態の顔画像による分析", 画像ラボ, 原稿 No. GA01-04, pp. 48-50, (1996. 4).