

論争支援のマルチモーダル実験システム

新田克己、長谷川修、秋葉友良、神畷敏弘、栗田多喜夫、速水悟、伊藤克巨

(電子技術総合研究所)

石塚満、土肥浩 (東京大学)

奥村学 (北陸先端科学技術大学院大学)

論争を支援する知識ベースシステムに、顔認識、表情合成、音声認識、音声合成、WWWブラウザなどのモジュールを結合したマルチモーダル実験システムを紹介する。このシステムは、原告(検察官)、被告側弁護士、裁判官という仮想的な3つのエージェントからなっている。ユーザは被告側弁護士に音声で指示を出して、検察官と法廷論争を行い、論争が終了すると裁判官が判決を下す。論争の状況に応じて、エージェントの表情が変化するので、ユーザはそれを見ながら論争の戦略をたてることができる。

An Experimental Multi-Modal Debate System

Katsumi Nitta, Osamu Hasegawa, Tomoyoshi Akiba, Toshihiro Kamishima,

Takio Kurita, Satoru Hayamizu, Katsunobu Itoh

(Electrotechnical Laboratory)

Mitsuru Ishizuka, Hiroshi Dohi (Tokyo University)

Manabu Okumura (Japan Advanced Institute for Science and Technology)

An Experimental Multi-Modal Debate System is a knowledge base system with user interfaces such as face recognition, face generation, speech recognition, speech generation and a WWW browser. This system consists of three agents - a prosecution, an attorney and a judge. The user helps the attorney to dispute with the prosecution. After the debate finishes, the judge makes the decision. According to the status of the debate, faces of agents changes. Therefore, the user understands the status, and can select the best move of the debate.

1 はじめに

推論機構、自然言語処理などの「記号による情報処理技術」や、画像認識、音声認識などの「パターン情報処理技術」が、開発されてきている。これらの多様な情報処理技術を統合することによって、相互に補い合い、より高度な情報処理機能が実現できるものと期待されている。

われわれは、分散環境のもとでのマルチモーダル情報処理（マルチモーダルネットワークエージェント）の研究を進めるためのプロトタイプとして、論争を行う推論モジュール、顔を認識するモジュール、表情を生成するモジュール、音声を認識するモジュール、WWW から情報を検索するモジュール、などを結合した実験システム Mr.Bengo を開発した。このシステムは、原告（検察官）、弁護士、裁判官という3つの仮想的なエージェントからなる。他人から訴えられたユーザは弁護士と相談しながら、原告に対抗して、法廷論争を繰り広げる機能を持つ。

2 システムの概要

MrBengo を起動すると、画面に3つの顔が現れる（図1）。これは、原告、被告側弁護士、裁判官の各エージェントを表す。



図 1: MrBengo の実行画面

このシステムは以下の4つの機能を有する。

- (1) 弁護士はテレビカメラに写ったユーザの顔を見て、誰であるかを認識し、その人の事件を思い出して内容を確認し、さらに補足質問を行う。
- (2) 原告は、「被告に一定の義務がある（または、原告に一定の権利がある）」ことを立証しようとして、論争を挑んでくる。ユーザは弁護士と相談しながら、反論を作って、原告につきつける。原告は、それに対してさらに反論をしてくる。
- (3) ユーザは必要に応じて、WWW の情報検索モジュールを起動し、WWW の中に関連する法令や判例があるかどうかを検索する。
- (4) 論争を終えると、画面には裁判官の顔が現れ、論争について、最終的な判断を下す。

これら(1)～(4)において、ユーザ、原告、弁護士、裁判官とのやりとりや、検索モジュールへの指示はすべて音声で行われる。

次章以降で MrBengo を構成するモジュールの概要を説明する。

3 論争モジュール

論争モジュールは、第5世代コンピュータプロジェクトで開発された法的推論システム New HELIC-II [5] を母体としている。このモジュールは、「論証生成」、「論証選択」、「論争支援」、の3つのサブモジュールと、「概念辞書」、「ルールベース（法令、判例、一般常識）」、「判断基準」、という3つの知識ベースからなっている。「判断基準」とは、「報道の自由」、「人権の保護」、「社会秩序の維持」などのように、ルールの優先関係を決めるときの基準となるものである。この基準間の優先関係を「価値観」と呼ぶ。価値観は人によって異なる。

原告と被告が論争を行う場合には、2つの推論モジュールが論争支援サブモジュールを介して、互いに論争情報を交換することによって論争が進行する。

以下にモジュールの内容を簡単に紹介する。

(i) 論証生成サブモジュール:

与えられた質問（ゴール）に対して、このゴールが成立するかどうかをルールを組み合わせで証明し、その論証（具象化されたルールの連鎖）を出力する機能を持つ。このサブモジュールは、さらに、生成された論証の各中間ゴールに対して、それらの否定が証明できるかどうかをチェックし、証明できるならば、もとの論証の反証として出力する。

(例) 論証 A { $r1 :: p(a) \leftarrow q(a), s(a)$, $r2 :: q(a) \leftarrow r(a)$, $f1 :: r(a)$, $f2 :: s(a)$ }.
反論 C { $r3 :: \neg q(a) \leftarrow t(a), v(a)$, $r4 :: t(a) \leftarrow u(a)$, $f3 :: u(a)$, $f4 :: v(a)$ }.

(ii) 論証選択サブモジュール:

論証と反論を比較し、どちらが優先するかを決定する機能を持つ。具体的には、価値観を参照して、論証と反論で対立するルール（上の例では $r2$ と $r3$ ）の優先関係を調べる。反論のルールの方が優先する ($r3 > r2$) ならば、反論 C がもとの論証 A より優先する ($C \gg A$)。ただし、原告と被告の価値観は異なるので、論証と反証の間の優先関係は、双方で異なる可能性がある。

(iii) 論争支援サブモジュール:

その当事者の価値観と、論争を行うための以下の5種類的手段と、これらを使って論争をする戦略アルゴリズムからなっている。（以下の論証手段において、括弧内は、ユーザが弁護士エージェントに命令するときの話し言葉の例である。）

- (1) **pose** : 生成した論証を相手に提示する。（「これを我々の論証としたい」）
- (2) **show_counter** : 相手から送られた論証の中で、反論すべき点（論点）を抽出し、それを新しいゴールとして、その論証を生成する。（「～について反論したい」）
- (3) **justify** : 自分の論証（反証）が相手の反証（論証）より優先することを主張するため、価値観を追加する。（「論証を正当化したい」）

- (4) **notify** : 自分の論証 (反証) が相手の反証 (論証) より優先することを主張する。 (「我々の論証の方が有利です。」)
- (5) **cancel** : 自分の論証 (反証) が相手の反証 (論証) より優先することが主張することができないので、論点を変更する。 (「論証を取り下げます」)

ユーザは弁護士エージェントに上記の命令をしながら、論争をコントロールする。一方、原告エージェントは、上記の論争手段を使った論争戦略のアルゴリズムを使って、論争をしかけてくる。

典型的な論争の経過は以下のようなものである。

- (step1) 原告が被告に論証を提示。
(例) 「被告は過失致死罪である。なぜなら、～だからである。」
- (step2) 被告は論点を抽出し、反証を生成して原告に提示。
(例) 「被告には過失はなかった。なぜなら泥酔していたからである。」
- (step3) 原告は判断基準を追加して、自分の論証の方が反証より強いことを主張。
(例) 「アルコール中毒の場合は、泥酔したこと自体に過失があるから、反論は成り立たない。」
- (step4) 被告は別の論点を選択し、反証を生成して原告に提示。
(例) 「原告にも過失があった。なぜなら、～だからである。」
- (step5) 原告は被告の反証に対して論点を抽出して、再反証を生成して被告に提示。
(例) 「原告には、過失はなかった。なぜなら、～だからである。」

.....

4 顔認識モジュール

顔認識モジュールは、テレビカメラに写った顔画像から誰かを特定する機能を提供する。

このモジュールでは、画像の特徴として高次局所自己相関特徴を用い、顔画像の例からの学習には、判別分析を用いた。

自己相関関数は平行移動に関して不変であることが知られている。その高次への拡張が、「高次自己相関関数」である。例えば、 N 次自己相関関数は、変位方向 $(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N)$ に対して、

$$x^N(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N) = \int f(\mathbf{r})f(\mathbf{r} + \mathbf{a}_1) \cdots f(\mathbf{r} + \mathbf{a}_N) d\mathbf{r} \quad (1)$$

で定義される。ただし、 $f(\mathbf{r})$ は、テレビカメラでとらえた認識・計測対象の画像である。

高次自己相関関数は、次数や変位方向 $(\mathbf{a}_1, \dots, \mathbf{a}_N)$ の取り方により、無数に考えられる。一般に画像中の近くの画素間の相関は高く、画素間の距離が大きくなるにつれて相関が低くなる。そこで、変位方向を参照点 \mathbf{r} の局所領域に限定し、遠くの画素との相関は無視するものとする。これを高次局所自己相関と呼ぶ。

高次局所自己相関に基づく初期特徴は、認識課題に依存しない一般的で基本的な特徴となっており、認識に必要な対象画像の情報は全体としてある程度抽出されていると考えられる。対象を識別するのに有効な特徴を構成するためには、さらに、これらの初期特徴を統合（総合）する必要がある。このためには、ニューラルネットを用いることも可能であるが、ここでは、最も簡単な方法として、初期特徴 $\{x_i | i = 1, \dots, M\}$ の線形結合

$$y_j = \sum_{i=1}^M a_{ij} x_i \quad (j = 1, \dots, N) \quad \mathbf{y} = A^T \mathbf{x} \quad (2)$$

によって、有効な特徴を抽出することを考える。ここで、 $\{a_{ij}\}$ は、結合係数である。また、 M は、初期特徴の個数であり、 A^T は行列 $A = [a_{ij}]$ の転置を表す。

今、初期特徴ベクトル \mathbf{x} の K クラスのサンプル集合 $C_k = \{\mathbf{x}\} (k = 1, \dots, K)$ が与えられているとする。この時、線形判別分析のための最適な係数行列 A は、固有値問題

$$\Sigma_B A = \Sigma_W A \Lambda, \quad A^T \Sigma_W A = I_N \quad (3)$$

の解として求まる。ここで、 N は判別空間の次元で、 $N \leq \min(K-1, m)$ である。また、 Σ_B および Σ_W は、それぞれ、クラス間共分散行列およびクラス内共分散行列である。

未知の画像が与えられた場合の識別は、例えば、その画像の初期特徴 \mathbf{x} から新特徴 \mathbf{y} を計算し、それに最も近いクラス平均ベクトル $\bar{\mathbf{y}}_k = A^T \bar{\mathbf{x}}_k$ を持つクラスに決定するなどにより実現できる。

5 表情合成モジュール

表情合成モジュールは、原告、被告弁護士、裁判官の3つのエージェントの表情を生成する。

本システムで用いたエージェントのモデルは、3種ともに3次元的に配置した約800個のパーティックス（点）と約1500個のポリゴン（微小三角平面）から成り、各ポリゴンを適切な色でレンダリングすることによって、「顔」としての外観を整えている。

本システムの3種のエージェントは、平常状態に加え、怒り、悲しみ、喜び、驚き、の各表情と、うなずき、瞬きの挙動の表出が可能となっている。

表情の表出は、エージェント毎に各表情のポリゴンのパターンを予め用意し、ある表情とこれから表出しようとする表情のパターンの差を5ステップで補間することにより行なっている。この操作は顔全体に対し一律に行なう以外にも、眉、まぶた、口に対して選択的に行なうことができ、例えば口だけを笑わせるなどの処理が可能である。

挙動の表出は、表情とは独立に、主として頭の回転運動を細かく制御することにより実現している。これにより、エージェントは笑顔や困惑の表情を表出しつつ、うなずくことができる。

基本的にエージェントの描画は、C++（実際にはg++）とOpenGL、GLUTによって行なっている。以下の *command*, *mface*, *agent*, *obj* は、それぞれ本モジュールに含まれるおける個々のプログラムの名称である。

以下に、エージェントの合成に至る過程の概要を述べる。基本的に顔画像合成モジュールとしての本プログラムは、サーバとして機能する。

- (1) *command*: ソケットを通じエージェントの描画要求を受けとる。
- (2) *mface*: 受けとった描画要求に基づいて全てのエージェントの挙動を管理し、*agent* に対してエージェントの起動順に描画要求を出す。
- (3) *agent*: *mface* から要求されたエージェントの挙動を個別に解析し、*obj* にそれらの描画要求を出す。
- (4) *obj*: 描画すべきエージェントの状態や、視点・光源位置の算出を行ない、この算出結果に基づいて、OpenGL に対して描画を要求する。
- (5) OpenGL: エージェントの描画を行う。
- (6) GLUT: OpenGL が描画すべきウィンドウの管理や、ポップアップメニューの提供などを行う。
- (7) X: 3 種のエージェントに各々固有のウィンドウを提供し、それらをモニタ上に表示する。

6 音声認識モジュール

音声認識モジュールは、「音声入力モジュール」と「音声認識管理モジュール」から構成される。

「音声入力モジュール」は、マイクロフォンからの音声信号を入力として、音声認識を行い、単語列を出力する。まず、音声信号から無音区間にはさまれた有音区間を切り出す。その音声区間にもっとも良く当てはまる音韻列を探し、認識結果とする。文脈自由文法で記述された文法と辞書によって可能な音韻列を指定し、認識の際の言語モデルとして利用する。認識結果は、この文法と辞書によって記述された可能な単語列のいずれかとなる。認識スコアとともに上位 N 個の単語列が、認識結果として出力される。

「音声認識管理モジュール」は、音声入力モジュールと、管理モジュールの橋渡しを行う。管理モジュールの発するユーザ・コマンドの要求に従って、音声入力モジュールに認識要求を発する。認識結果の単語列を受けとり、管理モジュールへの返答へと解釈を行う。解釈の過程には、単語列からコマンドへの変換を扱う自然言語処理と、直前にシステムが発話した内容のどれを指しているかを判定する文脈処理が含まれる。例えば、原告エージェントが次のように論証を発話したとする。

原告: 「... 19 歳の太郎が、契約をする時は、父の同意が必要である。...」

この後、弁護士エージェントに対して、ユーザが次の発話をしたとする。

ユーザ: 「同意が必要であるという点に反論してみてください。」

自然言語処理では、この発話からユーザが反論を行っている (コマンド 'show counter') と解釈する。さらに文脈処理では、「同意が必要である」が相手の主張のどの部分に当たるか (論証木の何番目であるか) を判断する。最終的には、入力を "show_counter 11" (推論木の 11 番目に対する反論を示せ) と解釈し、管理モジュールへ送る。

さらに音声認識管理モジュールでは、ユーザの判例検索要求を認識し、WWW 情報検索モジュールへの切替を行なう。ユーザは特定のキーワード (「過去の事例を検索したい」) を発することで、いつでも情報検索用 WWW ブラウザを呼び出すことができる。WWW ブラウザ上での情報検索にも、そのまま音声入力を利用する。

7 WWW ブラウザインタフェース

WWW ブラウザインタフェースは、音声を用いて、インターネット上の情報検索を行うモジュールである。

WWW ブラウザインタフェースは、C 言語で記述されたインタフェースライブラリと、外部プロセスから制御できるように改造された NCSA Mosaic で構成されている。インタフェースライブラリを含むアプリケーションプログラムと Mosaic は、二つの独立したプロセスとして実行される。これは、できるだけ元の Mosaic の機能に影響を与えないようにするためである。これにより改造された Mosaic は、アプリケーションプログラムとの接続に失敗した場合でも通常の Mosaic として実行できる。またネットワークに対応しており、アプリケーションプログラム、改造された Mosaic、およびその表示は、それぞれ異なるワークステーション上で実行することが可能である。

インタフェースライブラリには、必要な情報を含むリンクを選択するための 3 種類の関数が用意されている。

(1) アンカー文字列による指定

Mosaic 画面に現れるアンカー文字列を指定（発話）することにより、そのアンカーをマウスでクリックしたのと同じ効果が得られる。アンカー文字列から URL への変換は、改造した Mosaic 側で処理される。したがって異なるページ上で同じ文字列のアンカーが存在しても、それぞれ正しいリンクをたどることができる。

(2) インデックス番号による指定

画像や音声のように非文字列のアンカーを扱えることは HTML の優れた特徴の一つである。しかし、それらは文字列を持たないので、そのアンカーを音声で直接指定することはできない。そこで改造した Mosaic は、新しいページがオープンされる毎にそのページに含まれるアンカーの一覧表（アンカーリスト）をユーザに提示する。アンカーリストは、Mosaic 本体とは別のウインドウに表示される。それぞれのアンカー文字列にはインデックス番号が付けられている。ユーザはアンカー文字列を発話する代わりに、このインデックス番号を使用して任意のアンカーを選択することができる。(3) URL による指定

あらかじめ、オープンするページの URL がわかっている場合には、直接、URL を指定することができる。

8 管理モジュールと表示管理モジュール

「管理モジュール」は、各モジュールの間の通信を行ったり、エージェントを呼び出して起動させたりする機能を持つ。「表示管理モジュール」は、論争の状況に応じて、各エージェントの表情を指定したり、ユーザとの対話をスムーズに行うための会話文を音声出力したりするモジュールである。

9 おわりに

論争を行うマルチモーダル実験システム MrBengo の概要を紹介した。音声認識、音声合成、表情合成、顔認識、WWW ブラウザなどのインターフェースの結合により、従来、使い勝手の悪かった論争システムが非常に身近なものになった。

現在は、単一のワークステーションの上で3つのエージェントが動作しているが、これをネットワークを介して結合された分散環境における汎用のマルチモーダル情報処理システムに拡張することを検討している。

参考文献

- [1] 新田克己、長谷川修、秋葉友良、神畷敏弘、栗田多喜夫、速水悟、伊藤克亘、石塚満、土肥浩、奥村学. 論争支援のマルチモーダル実験システム. つくばソフトウェアシンポジウム '96.
- [2] 伊藤克亘, 速水悟, 田中穂積. 拡張 LR 構文解析法を用いた連続音声認識. 信学会技術報告, Vol.SP90, No.74, pp.49-56, 1990.
- [3] 相場徹, 堀田剛志, 本田岳夫, 奥村学. 音声対話文の意味解析に関する研究. 受託研究成果報告書, 1996.
- [4] Kurita T., Otsu N. and Sato T., "A face recognition method using higher order local autocorrelation and multivariate analysis," Proc. of 11th ICPR, The Hague, Vol.II, pp. 213-216, 1992.
- [5] Nitta K. et al., "New HELIC-II: A Software Tool for Legal Reasoning," *Proc. Int. Conf. on AI and Law*, 1995.
- [6] Nagao K. and Takeuchi A.:"Social Interaction: Multimodal Conversation with Social Agents" Proc. 12th AAAI 1992
- [7] 石塚満他:"新しいヒューマンインタフェースへ向けての並列コンピュータ (TN-VIT) 上のビジュアル・ソフトウェアエージェント (VSA)", 生産研究 (東京大学生産技術研究所発行), 44 巻 11 号, pp.7-15, 1992
- [8] O.Hasegawa et al.:"Active Agent Oriented Multimodal Interface System", Proc. AAAI International Joint Conference on Artificial Intelligence 95, Vol.1, pp.82-87 1995
- [9] 土肥, 石塚: 「WWW/Mosaic と結合した自然感の高い擬人化エージェントインタフェース」, 信学論 D-II, Vol.J79-D-II, No.4, pp.585-591, 1996.