

# 絵情報共有型会話エージェント Pictgent における 対話支援のための絵モデルの構築

上野 未貴      森 直樹      松本 啓之亮

大阪府立大学 工学研究科

ueno@ss.cs.osakafu-u.ac.jp, {mori, matsu}@cs.osakafu-u.ac.jp

## 1 はじめに

近年、計算機と人間とのコミュニケーションの実現を目指して、コンピュータによる人間の意図理解が注目されている。代表的な意図表現手段であるテキストデータに関しては、大規模データからの統計的な情報抽出が進んでいるが、一方で、対話において、テキストデータのみからの意図の抽出は極めて困難である。

また、近年のソーシャルメディアの発展を背景として、人工会話システムに注目が集まっている。筆者らはその中で特に人工無脳に着目し、ユーザと絵情報を共有して会話を実現する絵情報共有型会話エージェント (Picture Information Shared Conversation Agent: Pictgent) [1] を提案してきた。

状況依存の対話システムとしては、Winograd らの積み木世界の対話的操作の SHRDLU [2] が成功例の先駆けとして知られている。しかし、SHRDLU では共有している世界の自由度を大幅に制限することによってユーザの意図を限定し、対話を進めるため、計算機への親近感や共感という点では不十分である。一方で、人工無脳と呼ばれるシステムでは、ユーザの意図を厳密に解釈することは放棄し、表層的な部分で「人らしい」と錯覚できる対話を楽しむことを目的として、アプリケーションとしては一定の成功を収めている。しかしながら、ユーザの興味を維持するのが難しいという問題がある。

Pictgent は絵によってシーンを表すことにより、自然な状況限定や知識共有をした上で対話するシステムであり、SHRDLU と比較した場合には、共有する世界の自由度は大幅に増加している。また、人工無脳と比較した場合には、絵により事前に可能な状態遷移をシナリオとして規定するために、会話の発散や無目的感から生じる興味の喪失を防止することが期待できる。

本研究では、Pictgent で対話を実現する場合の絵モデルの策定や、またどのようにそれが対話の支援を果

たすかということシステムとの機能と関連づけて述べる。また、実際の人からの入力情報を定量化し、状態遷移に用いるためのデータ収集実験で用いる絵モデルの構成について示す。

以下に本研究の構成を示す。第 2 で関連研究を述べて本研究の位置づけを明確にし、第 3 章で Pictgent の概要および、絵モデルについて述べる。第 4 章で今後の課題を述べてまとめる。

## 2 関連研究と本研究の位置づけ

対話システムの研究には、自然言語処理だけではなく、言語学、進化言語学、認知心理学 [3]、心理学、哲学など様々な領域からのアプローチが重要となる。以下では本研究と関連のある従来研究と本研究の位置づけについて示す。

### 2.1 絵の一意性と多義性

パーツ分けされた絵についての情報付与の研究として、伊藤ら [4] のピクトグラムへのセマンティクスの導入手法が提案されている。この手法では、ピクトグラムを組合わせて、文を作成したり、逆に、文からピクトグラムを組合わせた絵に変換することができる。同研究では、言葉を理解することが難しい人でも絵を使うことによって防災情報など重要な行動が直感的に理解できることを目的としているため、多義性を含まない絵を作ることが重要視されていると考えられる。また、言語に対応した絵を表示することにより語義の曖昧性を解消する研究も報告されている。前述の両手法とも、絵と文を一意に相互変換することを目的としているため、絵を見る人によって多義性が生じることは好ましくない。これらの手法は一般的に共通認識が得られている概念や行動に記号的なラベルの意味で絵を 1 対 1 で対応させていると考えられる。

一方、本研究では、一つの絵とユーザの想起内容が1対多であることを仮定し、想起内容の多様性を用いて、一つの絵に複数の意図を対応づける。絵の解釈の多様性を心理学に応用したものに主題統覚検査 (Theoretical Application Test:TAT) [5] がある。Pictgent では、絵をユーザがどのように理解するか、といったユーザの絵理解の手順を対話中に集めることも可能である。

## 2.2 絵へのアノテーションとタグ構造

絵モデルの構築という静画への情報付与に関連して、山本ら [6] によって、動画をシーンごとに分け、Web上のユーザがタグ付けするシステムも提案されている。

また、Ahn[7] によって提案されている EPS ゲームは、二人のユーザが一つの画像に同じタグを付けると、スコアが得られるゲームであり、人手によるタグ付けのコスト低下と一般化の仕組みが提案され、Google Image Labeler で実際に技術が用いられている。

しかし、これらは画像に対して平坦なタグ付けをするものであり、構造を持たない。また、タグをそのまま対話に用いることは難しい。本研究では、対話に関わるタグの構造の精査も目的とする。

## 2.3 テキストデータからの意図理解

白木ら [8] はホテル予約というタスクを設定した音声認識対話システムにおける意図理解のため、文の一部を変数化し一般化した後、文をベクトルに直し、統計的に数種類の意図に分類する手法を述べている。

絵と文が既に対応づけられた題材としては、漫画が知られる。徳島ら [9] は、漫画の台詞を抜き出し、ラベルを付して、対となる台詞の関係性からの意図理解について言及している。

前者は、文のベクトル化に際して必要となる処理が述べられているが、考慮されているのは感情的な意図の理解ではない。後者は絵と文の対応づけのある題材を使用しているものの、分析には特にその関連性を意識せず、テキストデータを中心としている。

## 2.4 本研究の位置づけ

上記で述べた研究では、シーン間の意図を考慮したり、アノテーションする場合に、絵の中の主体客関係を示すものはない。また、動画視聴などの受動的なコンテンツでは、ユーザからの情報付加を期待するしかなく、能動的な情報付与は難しい。しかしなが

ら、Pictgent では、絵の内容に関する文を示し、対話を通じてユーザに問い返すことができるため、シーンへのアノテーションにも応用可能である。

我々は人間が日常であるシーンにおいて対話する際、語用論が示すように、意味どおりではない状況を判断するメカニズムを考慮して対話システムを構築することが重要だと考えている。そこで、テキストの行間を読む意図理解に相当するものを、絵でシーンとして表して離散的につないでシナリオを作成することにより、対話を支援する絵モデルを構築する。この複数の絵モデルにより構成されるシナリオは、分岐を許す自由度の高いものであり、ユーザの興味を十分に満たすことができる。また、シナリオから離れた会話も許容するため、シナリオに沿った入力を強要されていると感じるようなユーザの心理的負担の軽減も実現している。

また、Pictgent における意図理解という観点からは、絵モデルを用いることによって、ユーザの入力と絵の内容に関連があると仮定できる場合は、不自然な形ではなくユーザの意図を限定できる。絵モデルの遷移は、現在の絵に留まるか否か、次の絵モデルに遷移すべきか否かのようにユーザの入力を2値で判定すれば十分な場合がほとんどである。

Pictgent の適用対象は限定しないが、低年齢層のユーザであっても楽しんで使うことができるアプリケーションとして成立するように構築する。また、各種対話処理や絵情報に関する要素技術のテストベットのな利用も視野に入れて、モジュール性の高い柔軟な対話フレームワークとして本システムを位置づける。

# 3 Pictgent の概要

## 3.1 基本方針

Pictgent とはシステム側で用意した絵モデルについての情報をユーザと共有しながら会話が可能な対話システムである。また Pictgent が扱う絵モデルでは、ソシオグラムとオブジェクト図を融合している。ソシオグラムとは、小集団の人間関係や集団構造を図表で表示したものである。人間の属性だけでなく、動作の対象関係や人間の社会的関係も記述し、より詳細に絵の内容を扱うことを目的とする。

以下に Pictgent を構成する各モジュールを示す。

絵モジュール Pictgnet の中心的な要素である、絵を示すモジュールである。本モジュールは絵の内容を表すモデルを XML 形式で持つ。

```

<character>
  <base id="1" type="人" name="はな" sex="女" age="6" />
  <scene>
    <position>右</position>
    <expression>不思議そうな</expression>
    <emotion target="2">ドキドキした</emotion>
    <action name="見る" target="2" />
    <state>
      <physical></physical>
      <mental></mental>
    </state>
  </scene>
  <relation name="-" role="-" target="2" />
</character>
(中略)
<layer>
  <part id="1" visible="true">
  <part id="2" visible="false">
  <part id="3" visible="true">
  <part id="4" visible="true">
  <part id="5" visible="true">
</layer>

```

図 1: 絵モデルの人物とパーツに関する XML

表 1: 絵モデルで用いる主要タグ

タグ名	説明
<character>	人物オブジェクト.
<base>	登場人物に付随し, 場面転換で一般的に変わらない特徴.
<scene>	絵・場面に依りての情報.
<action>	動作. 対象を必ず付す.
<state>	オブジェクトの内部状態や付帯状況. 対象をもたない.
<relation>	他のキャラクタとの社会的・静的な関係性. (0..*)
<part>	配置可能な絵パーツと表示状態. (0..*)

表 2: 絵パーツ DB の一部

partid	sceneid	imagename	filepath
1	1	コート, 上着	coat1-1.png
1	2	コート, 上着	coat1-2.png

イメージ名はカンマ区切りで複数保持可能

シナリオ進行モジュール ユーザの各種情報を内部状態として保持し, また, ユーザの入力に応じて適切なシナリオの遷移を管理するモジュールである. シナリオは XML 形式で持ち, 遷移条件と遷移先と出力内容を記述している.

会話モジュール 実際の入力を受けて出力を生成するまでの内部処理をする会話ユニットを切り替えて管理し, 実行するモジュールである. Pictgent には現在大きく分けて雑談モードとシナリオモードと呼ばれる二つのモードが存在する. この二つのモードではそれぞれ人工無脳型とシナリオ型の会話ユニットを個別に用いる.

### 3.2 絵モデル

本研究では, 絵モデルをオブジェクト指向モデリングに基づき構成する. 今回は, 特に主要な登場人物のみをオブジェクトとして定義したが, シナリオに直接的には関わらない人物や無生物をオブジェクトとすることも可能である. 絵モデルは, XML 形式で保存さ



図 2: 絵 1-ベース 図 3: 絵 1-修飾後 図 4: パーツ



図 5: 絵 2-ベース 図 6: 絵 2-修飾後

```

<unit_phase>
  <pict_text>ふわあ、さむいよ。ぼくのおうちのやねをかえしてよ！</pict_text>
  <out>あれあれ、ゆきにあいたあなからかえるくんがでてきたよ。おうちのやねってなんだろうね？</out>
  <out>ねえ、<personname:1>が いまてにもっているものはなに？</out>
  <con>null</con>
  <yes next="1">そうだね。おはなはとうみんちゅうのかえるくんのやねだったんだね。</yes>
  <no next="2">うーん。かえるくんのしたからごそごそおとがするよ？</no>
</unit_phase>

```

図 7: シナリオを表す XML の例

表 3: シナリオの XML で用いる主要タグ

タグ名	説明
<unit_phase>	ユニットフェーズオブジェクト.
<uid>	ユニットフェーズの ID.
<pict_text>	絵内のキャラクタに関する出力. (0..*)
<out>	ユニットフェーズ開始時のボットの出力. (1..*)
<con>	ユニットフェーズ開始時の条件及び処理方法.
<yes>	ユーザの入力が正のときの出力, 属性として同一シナリオ内での遷移先のユニットフェーズの ID を持つことがある. (1..*)
<no>	ユーザの入力が負のときの出力, 属性として同一シナリオ内での遷移先のユニットフェーズの ID を持つことがある. (1..*)

れ, 必要に応じてモデルの編集も可能とする. 図 1 に図 3 に対応する絵モデルの XML のうち人物情報と絵パーツに関する部分を抜粋した例を示す. また, 表 1 に主要なタグを示す.

シナリオや対話内容によって, 絵も遷移する. 図 2 ~ 図 6 に絵とパーツの例を示す. また, シーン間での大きな変化だけではなく, 絵内において, パーツの配置によってベースとなる絵を修飾する小さな変化 (図 2 図 3, 図 5 図 6) も可能である. その際に生じた変化が, 例えば服を変更するような静的な変化の場合

には、絵間の遷移があった際に次の絵モデルの状態に影響を及ぼすことがある(図3 図6)。しかし、同じコートのパーツであっても、絵によって形状が変化するため、絵パーツの id と読み込む絵モデルの id を複合キーとして、絵パーツ DB から、適切な絵パーツのパスを読み出して配置する。表2に絵パーツ DB のテーブル構造の一部を示す。また対話による問い返しと誘導により、ユーザの入力と絵のどの要素が対応しているかを取得する。

図7に図3に対応するシナリオ XML 中の一つの単位である UnitPhase の例を、表3にシナリオ XML の主要なタグについて示す。

### 3.3 絵モデルの XML の利用手法

絵モデルとして記述した XML の情報は XML パーサーを用いてインスタンスにマッピングする。また、マッピングした要素によってテーブルや、シナリオ XML 内において、要素に対応したタグが置換される。絵モデルの XML とシナリオ XML を組み合わせ、図7の <personname:1> を図1の name が示す「はな」で置換するように、再利用可能部分をタグでの置換で済ませることにより、新たにシナリオを書き起こすコストが減少する。

用意した絵とシナリオを使って、絵モデルによる対話支援が実際にユーザにどのように意図理解を促すか調べるため、幼稚園～高校生を対象とした科学イベントで実験する。その際に、発話と作成された絵モデルの関係をログに取得して分析する。なお、図7の1つ目の <out> タグが示すような、絵だけで場面を理解できたか、2つ目の <out> タグが示すように、システムの補助発言によって場面を理解できたか、シーンごとの発言内容の頻度などの結果を発表時に述べる。

## 4 まとめと今後の課題

本研究では、Pictgent の関連研究との差異について詳述することにより位置づけを明確にした。また、絵モデルの構築方針について述べた。今後の課題として、絵モデルやシナリオ XML の情報を増やした上で、実験を重ねてログを分析することにより、対話に即して意図を抽出できるよう構成を精査することが挙げられる。また、現在 Pictgent へ実装済みであり、Twitter 上でもプロトタイプ [10] を一部稼働させている言語からの絵を生成する機能を拡張し、絵モデルの自動生成のアルゴリズムを検討する。

なお、本研究は、日本学術振興会科学研究補助金基盤研究(C)(課題番号 22500208)の補助を得て行われたものである。

## 参考文献

- [1] Miki Ueno, Naoki Mori, Keinosuke Matsumoto : Novel Chatterbot System Utilizing Web Information for Estimating Current User Interests, Proc. of the 6th IEEE International Conference on IDAACS, Vol.2, 656-659, ISBN : 978-1-4577-1424-5 (2011)
- [2] Terry Winograd : Procedures as a Representation for Data in a Computer Program for Understanding Natural Language , Cognitive Psychology Vol. 3 No 1 (1972)
- [3] 道又爾, 北崎充晃, 大久保衛亜 : 認知心理学: 知のアーキテクチャを探る, 有斐閣 (2003)
- [4] Kazunari Ito, Motohiro Matsuda, Martin J. Durst and Koiti Hasida : SVG Pictograms with Natural Language Based and Semantic Information, SVG Open (2007)
- [5] Morgan, C.D., Murray,H.A. : A method of investigating fantasies. Archives of Neurology and Psychiatry, 34, 289-306 (1935)
- [6] 山本大介, 増田 智樹, 大平 茂輝, 長尾 確 : タグクラウド共有に基づく協調的映像アノテーション, 人工知能学会論文誌, Vol.25, No.2, 243-251 (2010)
- [7] Luis von Ahn, Laura Dabbish : Labeling images with a computer game, CHI'2004, 319-326 (2004)
- [8] 白木 将幸, 伊藤 敏彦, 甲斐 充彦, 中谷 広正 : 自然発話文における統計的な意図理解手法の検討, 情報処理学会研究報告. SLP, 音声言語情報処理 2004(15), 69-74 (2004)
- [9] 徳久 雅人, 村上 仁一, 池原 悟 : テキスト対話コーパスからの発話対と情緒の分析, 電子情報通信学会技術研究報告.TL, 思考と言語 108(50), 41-46 (2008)
- [10] @cotopaint - ことべいんと : <http://zope.ss.cs.osakafu-u.ac.jp/Members/ueno/cotopaint/>