

# 絵内オブジェクトの状態遷移特性を利用した絵モデルの構築

## The Picture Model by means of Transition Information of Objects in Picture

上野 未貴<sup>1\*</sup> 森 直樹<sup>1</sup> 松本 啓之亮<sup>1</sup>  
Miki Ueno<sup>1</sup> Naoki Mori<sup>1</sup> Keinosuke Matsumoto<sup>1</sup>

<sup>1</sup> 大阪府立大学 工学研究科

<sup>1</sup> College of Engineering, Osaka Prefecture University

**Abstract:** We have proposed conversation system sharing picture information with users called Pictgent and defined the picture model as XML format. Though the picture model is essential part of Pictgent, there are problems to define complete picture model because of high complexity of user inputs. In this paper, we proposed a picture model based on transition information of objects in pictures of four-scene comics.

## 1 はじめに

近年、計算機と人とのコミュニケーションに関して多くの研究がなされている。両者間のコミュニケーションにおいて最も重要であると考えられる自然言語処理の分野では、大規模なデータやコーパスを使ったテキスト分類などの研究が盛んだが、文章の意図理解はデータが豊富になった現在でも困難である。そこで、筆者らは、コミュニケーション時に絵を提示することによって、対話における状況を緩やかに設定し自然な対話を実現可能な絵情報共有型会話エージェント (Picture Information Shared Conversation Agent: Pictgent) を提案している。

絵は幼児でも理解でき、自然言語ほど地域性が強くないといった特徴があり、共通の絵について対話することによって、ユーザが戸惑うことなく対話を進められるという利点がある。Pictgent では、絵と絵に付加した情報を絵モデルとして定義し、対話時に利用したり、計算機で絵を扱う必要があるユーザに絵コーパスとして絵モデルを提供可能な環境を整備することを研究の目的の一つとしている。しかしながら、具体的にどのような形式で絵モデルを構築すれば良いかについては十分な検討がされておらず、大きな問題となっている。

画像認識の分野では、写真中の人物の識別や、物体の認識の研究が進んでいるが、絵の中で何が起きているかという状況の認識や、物語を聞いたときの話の流れを計算機に自動で解釈させるために用いることはできない。パーツ分けされた絵についての情報付与の研究としては、伊藤ら [1] のピクトグラムへのセマンティクスの導入の研究が知られるが、これは言語を理解する人が困難な人への絵を記号的な伝達手段と考えたも

のであり、感性的な情報には着目されていない。また、スクリプト構造や意味ネットワークなど、日常の話の流れや想起しやすい単語の連結など、人工知能の領域でいくつかの知識表現方法が提案されているが、計算機で絵の内容の表現方法は十分に検討されてこなかった。

そこで、本研究では計算機に絵を理解させるために必要な絵モデルの構築方法について考察する。現段階ではすべての絵に対して汎用的に用いることができる絵モデルの構築は困難であると考え、ユーザとの対話時における応用の容易さおよび意図のある絵であることに重点を置き、4コマ漫画を例にとり、内容理解のために必要な絵モデルを構築するための情報と構造について検討する。また、実際に構築した絵モデルを用いて、一般的な絵モデルに必要な要素および絵モデル構築時に困難となる事項について考察する。

## 2 絵の情報の付加

本章では、今回作成した絵の内容の具体例を示して、特徴的な要素について説明を加え、必要な要素を検討する。図1に情報を付加する対象となる4コマ漫画を示す。図2に作成した絵の情報を表すXMLを示す。表1に今回用いたタグの説明を示す。

### 2.1 絵モデル

本研究では、絵を以下に従ってモデル化する。

#### 2.1.1 絵内のオブジェクト

まず、絵内の各構成要素をオブジェクトと呼ぶ。基本的に絵を構成する要素はすべてオブジェクトであり、

\*連絡先: 大阪府立大学 工学研究科  
〒599-8531 堺市中区学園町1番1号  
E-mail: ueno@ss.osakafu-u.ac.jp

背景や吹き出しもオブジェクトとして扱う。各オブジェクトは名称や必要があれば外部から内容を表す情報を取得することができる。また、上にある、重なっている、包含している等のオブジェクト間の物理的関係と絵内におけるサイズと位置も保持する。また、オブジェクト間の has-a 関係も記録する。

### 2.1.2 オブジェクト間の非物理的関係

絵に描かれているオブジェクトは物理的関係以外にも多くの関係を持っている。例えば、社会的関係、感情的関係等がこれに相当する。また、4コマ漫画のように連続する絵の場合には、各コマ間のオブジェクトに時系列的な因果関係が存在する場合がある。例えば、卵から雞が出てくる、氷が溶ける、叩かれた結果コブができる等である。このような非物理的関係についても記述する。特に感情については対象を示した上で、ベクトルを用いて定量的に示す。

### 2.1.3 オブジェクトの状態変化分布

2.1.1 節、2.1.3 節により絵内のオブジェクトについての基本的な内容は表すことができる。しかしながら、それだけでは対話で利用するには不十分である。椅子が描かれていたときに「椅子とは人が座るための家具であり一般に四脚である」といった情報は重要であるが、それ自体を対話で直接的に用いるのは不自然なためである。一般に4コマ漫画に描かれているオブジェクトはストーリーに関わることが多い。例えば椅子があった場合には「椅子に座っていた人が転ぶかもしれない」とか「悪戯で誰かが椅子を後ろから引くかもしれない」といった情報の方が対話に役立つと考えられる。ストーリーが存在する絵内のオブジェクトは特徴的な変化をすることが多い。そこで既存の絵からそこに登場したオブジェクトのストーリー内での変化を調べ、その結果をデータベース化する。これによって得られる各オブジェクトの状態変化分布は人が絵を見た場合に想起する内容を推測したり、自然な対話を実現するために有用であると考えられる。

### 2.1.4 今後検討を要する項目

2.1.1 節については、表1で示した種類の1が、2.1.2 節には、2が該当する。2.1.3 節については今回の具体例一つでは不十分であるが、データベース化して共通情報が増えることで実現できると考える。

なお、has-a についてはオブジェクト指向モデリングの分野で議論がなされているように、どの粒度で物体を捉えるかが難しいが、ストーリーにおいて詳細な情報を

参照する可能性がある場合には、has-a の属性を使って、各タグに情報を付加することがある。また、emotion だけで表すことが難しいと思われる感情を mental に記述しているが、emotion のベクトルにすべてを統合できるかどうかについても、検討の必要がある。

## 2.2 対象とする絵

絵には動画と静止画があるが、今回は静止画に焦点を当てる。さらに連続した静止画によって、時間変化をする4コマ漫画を用いる。4コマ漫画では一般に複雑な描写は少なく、話の流れに関係する物体のみが描かれることから、全体としての情報量は少ないが、絵内の各オブジェクトが持つ情報量は大きいという利点がある。そのため、話の流れを決定する物体を限定しやすい。周りの不要な情報まで映す写真に比べて、人間が意図して情報を限定しているため、話や物体の役割が明確であり、今回の目的に適している。

本研究では、計算機において絵の内容および絵で表される話の流れをいかに構造化するかが重要であるため、絵が主体となり話の流れを表現している、話の流れが一般的に共感できる、という二つの観点を考慮して一つの4コマ漫画を具体例として選んだ。

上記に関して、これら二つの観点到にそぐわない例をそれぞれ挙げる。一つ目の観点到に反するものの例として、絵を補助的に用い、台詞に依存して話を表現しているもの、および、言い間違いや掛け言葉など言語そのものの特徴を話の主体としているものが挙げられる。4コマ漫画の特徴として、絵と共に台詞が描かれているが、絵と台詞の比重は、話によって異なる。

二つ目の観点到に反するものの例として、現実で起こり得ない突飛な展開を持つ話が挙げられる。日常では起こり得ないが、漫画では典型的な表現として確立されていることもある。

## 2.3 4コマ漫画の基本情報

以下では、XML の情報をストーリー全体、コマ、オブジェクト、という粒度の異なる単位で捉えた時の、各単位について、タグの説明だけでは不十分と思われる特徴的な情報について詳細に述べる。

ストーリー全体の構成として、ストーリーを一意に表す id 番号を付加する。その後、ストーリーで登場する人物や無生物などのオブジェクトの基本情報を記述し、各コマごとの情報を記述する。各コマには、登場するオブジェクトのコマに依存する情報を記述する。また、コマ内でのオブジェクト同士の関係を記述する。

以下にオブジェクトが持つべき情報を詳細に述べる。



図 1: 花火対決

(©Beeworks 出典: <http://namepara.com/info/dx/609.html>)

表 1: 今回作成した絵モデルで用いたタグ

タグ名	種類	説明
<story>		ストーリーの情報。ストーリーの id, タイトル, オチとなるコマの id, コマのサイズと単位を持つ。
<coma>		各コマの情報。
<base>	1	場面転換で一般的に変わらない特徴。
<character>		意思を持った登場人物。擬人化されている場合は, 見た目が人間以外の生物と無生物の場合も含む。見た目が人間以外の場合は, 見た目の情報を要素で示す。
<object>		登場人物意外で, 背景ではない物体。
<position> 1		オブジェクトの重心点の x 座標と y 座標およびコマを手前から見たときのコマ内での奥行きに関する順序。
<size>	1	参照とする id の登場人物を 1 とするおおよその大きさの割合。
<expression>	2	表情。
<physical>	2	身体的な状態。e.g.: 病気である, 怪我をしている。
<mental>	2	精神的な状態。e.g.: 迷っている, 落ち込んでいる。
<emotion>	2	感情。基本の 6 つの感情 [2] である「恐れ」「怒り」「嫌悪」「悲しみ」「驚き」「喜び」の各要素を 0 から 1 の値をとる実数値ベクトルで表す。
<action>	2	動作。対象を必ず付す。
<state>	2	オブジェクトの内部状態や付帯状況。対象をもたない。
<think>	2	登場人物の考え。
<icon>		漫符の意味。
<onomatopoeia>		絵中のオノマトペ。
<texticon>		絵中の書き言葉。
<text>		台詞。
<event>	2	話に直接的に関わる事象。事象が意図的なものかどうかを要素にもつ。
<relation>	2	登場人物間の社会的・静的な関係性。
<location>		場所。
<time>		時間。
<season>		季節。

### 2.3.1 登場人物, 生物と無生物について

背景以外で, 話の中で重要であると考えられる物体をオブジェクトとする。オブジェクトは自ら意思を持って動作できる自律的なものと, それ以外に大きく二つ

```
<story id="1" title="花火対決" punch-line="2, 4"
coma-height="190" coma-width="270" unit="px">
<character>
<base id="1" looks="なめこ" name="なめ
こ" sex="unknown" age="unknown"
reality="false" coma="1, 2, 3, 4" />
<relation name="上司-部下" role="部下" target="2" />
</character>
<character>
<base id="2" class="人" name="小沢里奈" sex="
女" age="unknown" reality="false" coma="1, 2, 3, 4" />
<relation name="上司-部下" role="上司" target="1" />
</character>
<object>
<base id="3" class="花火" kind="線香花火" name="線香花
火" reality="true" coma="1, 2" />
</object>
<object>
<base id="4" class="花火" kind="線香花火" name="線香花
火" reality="true" coma="1, 2, 3, 4" />
</object>
<object>
<base id="5" class="花火" kind="線香花火" name="線香花
火" reality="true" coma="3, 4" />
</object>
... (中略 1 ~ 3 コマ) ...
<coma id="4">
<character id="1">
<position x="34" y="125" z-rank="1"/>
<size ref="1" height="1" width="1"/>
<expression>陰った</expression>
<physical>null</physical>
<mental>ショックを受けている</mental>
<emotion>(0, 0, 0, 1, 0.4, 0)</emotion>
<action target="3">落とした</action>
<action target="3">[link ref="1" has-a="手"] を滑らせた
</action>
<state>立つ</state>
<think>[link ref="4" has-a="火薬部"] を落としたい。
</think>
</object>
<character id="2">
<position x="210" y="104" z="3"/>
<size ref="1" height="1.2" width="0.9"/>
<expression>無表情</expression>
<physical>null</physical>
<mental>null</mental>
<emotion>(0, 0, 0.2, 0.2, 0.1, 0)</emotion>
<action target="4">持つ</action>
<state>立つ</state>
<state>服を着ている</state>
<think>[link ref="1"] は運が悪い。</think>
</object>
<object id="4">
<position x="186" y="145" z-rank="1"/>
<size ref="1" height="0.25" width="1"/>
<state has-a="火薬部">燃える</state>
<state has-a="火薬部">火花を散らす</state>
</object>
<object id="5">
<position x="94" y="170" z-rank="3"/>
<size ref="1" height="0.25" width="1"/>
<state has-a="火薬部">が燃える</state>
<state has-a="火薬部">が火花を散らす</state>
<state>落ちる</state>
</object>
<icon owner="1">ショック</icon>
<icon owner="2">呆れる</icon>
<icon owner="1">滑る</icon>
<onomatopoeia owner="1">つるつ</onomatopoeia>
<event intentional="false">[link ref="5"] が落ちる</event>
<location background="false">unknown</location>
<time>unknown</time>
<season>unknown</season>
</coma>
</story>
```

図 2: 今回作成した絵モデルの XML

に分けることができる。提案手法ではこれらをそれぞれ, character と object という二つのタグで表す。また, 絵では実在しないものが描かれることもあるため, 実在するかを属性として示す。

### 2.3.2 大きさ

4コマ漫画では、コマの中で背景が描かれないことが多く、オブジェクトの大きさを推定することは難しい。そこで、ストーリーの中で基準とするオブジェクトを一つ選び、オブジェクトとの相対的な大きさを記述する。基準とするオブジェクトはコマで同時に描かれる他のオブジェクトの種数の合計が最も大きいものとする。ストーリーの中で、a, b, c という3つのオブジェクトがあるときに、1, 2コマ目でaとb, 2, 4コマ目でaとcが出現する場合、基準とするオブジェクトはaとなる。ただし、基準とするオブジェクトは登場人物の中から選ぶこととする。

また、話に大きさが関わってくる場合は、任意のオブジェクトの大きさに特徴があると考えられる。オブジェクトが属するクラスの中で、高さもしくは幅が特に大きいもしくは小さいと思われる時には、オブジェクトの属性として記述する。

### 2.3.3 位置

4コマ漫画では、オブジェクト全体が描かれないことも多い。しかしながら、人間は既知の物体であれば全体像を思い浮かべ、重心を推測することができる。絵は3次元世界を2次元で描くため、空間の描き方によっては、z軸の特定は難しい。そのため、x, y座標についてはコマの左上隅を原点、右下隅をコマのx, y両座標の最大値とし、z軸については、画面手前を最小としたコマ内に登場するオブジェクトの順序で表す。

## 2.4 漫画特有の情報

漫符という人物の感情やオブジェクトの動作などを静止画で表す際の補助として用いられる漫画特有の記号がある。また、描画の補助としてオノマトペが手書き文字として用いられることも多い。図1の例の4コマ目においては、落下を表す「つるっ」というオノマトペや、ショックを受けた感情を表す漫符が用いられている。これらは登場人物や無生物とは異なるタグで表す。これらの記号や書き文字に着目することで、コマの中での状態変化を推定できる可能性がある。

## 3 応用

本研究で作成した絵モデルは工学的に利用することを前提としており、応用先として、筆者らが提案しているPictgentという対話システムで用いることを考えている。Pictgentとは、絵を提示して、その内容についてユーザと対話をするシステムである。絵を用いるこ



図 3: Pictgent の GUI の概観

とにより、言語だけの対話では得られない絵に対する感性的な入力を得ることや、自然に状況を限定して対話を実現することが大きな特徴である。Pictgentの適用対象は限定しないが、低年齢層のユーザでも楽しんで使用できるアプリケーションとして成立するように構築している。図3にPictgentのGUIを示す。

Pictgentにおいては、複数の絵モデルを参照することで、エージェントがオブジェクト構成の類似した話の情報から次の状態を予測した文を発言することなどに役立つと考えている。

## 4 まとめと今後の課題

本研究では、工学的に用いる絵の情報付加方法の策定を目指し、具体例を挙げて問題を提起した。複数の絵モデルを比較検討し、共通要素と差異を抽出することを今後の課題とする。動作実験の例は発表時に示す。

## 謝辞

本研究は一部、日本学術振興会科学研究補助金基盤研究(C) (課題番号 22500208) および第9回リバネス研究費 DeNA 賞の補助を得て行われた。また、絵の引用許可を下さった株式会社ビーワークスに謝意を示す。

## 参考文献

- [1] Kazunari Ito, Motohiro Matsuda, Martin J. Durst and Koiti Hasida: SVG Pictograms with Natural Language Based and Semantic Information, SVG Open (2007)
- [2] Antonio R. Damasio: The Feeling of What Happens: Body and Emotion in the Making of Consciousness, Harcourt Brace and Co (1999)