

絵の時系列的状態変化分布に基づく絵モデルの構築

The Picture Model based on the Distribution of Picture State Transition

上野 未貴*¹ 森 直樹*¹ 松本 啓之亮*¹
 Miki Ueno Naoki Mori Keinosuke Matsumoto

*¹大阪府立大学, 工学研究科
 Graduate school of Engineering

Recently, picture understanding by computer has become important. Although humans can understand situations in pictures easily, there are few research have been reported about picture understanding like human by computer.

In this paper, we proposed the novel method which estimates the future variation of story and objects in picture in order to construct the suitable picture model for computers. In our model, we categorized story into 12 patterns based on punch line and analyzed the distribution of transition information of objects in pictures. The computational experiments are carried out with pictures of four-scene comics as examples.

1. はじめに

人間の知的活動の産物を計算機に理解させるという課題は、人工知能研究の大きな目的の一つである。人間の知的活動は多様な表現形態をとるため、計算機で扱う場合には、それぞれについて考察する必要がある。自然言語は重要な表現手法であり、比較的計算機で扱いやすいため多くの研究がなされている。

一方で、絵画や音楽など抽象性が高い芸術的分野は計算機による扱いが難しいとされてきた。例えば、絵であれば画像処理によって内容をある程度認識することは可能であるが、その絵に込められた心情や意味を読み取る方法については十分には研究されていない。しかしながら、「百聞は一見にしかず」という諺にもあるように、絵は状況によっては非常に強力なコミュニケーション手段となり得る。

以上の点を背景として、筆者らは、コミュニケーション時に絵を提示することで、対話の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) [Ueno 2012] を提案している。Pictgent では絵と絵の内容を表す情報を絵モデルとして定義し、対話時に利用しているが、絵モデルの完備な定義はなされていない。また、従来研究としても、計算機での絵の内容の表現方法は十分に検討されてこなかった [新井 2012]。

そこで、本研究では絵の時系列変化を統計的な分布として扱いデータベース化する手法を提案し、より汎用性の高い絵モデルの構築を目指す。具体的には4コマ漫画にはストーリー展開とオブジェクトの時系列変化に特定のパターンが存在するという仮定の下、実際に時系列変化を調べることによって提案手法の有効性を示す。

2. 計算機による絵の扱い

本章では、本研究における絵の扱いについて述べる。

2.1 これまでの絵の扱い

計算機による絵の扱いは画像処理が中心であり、写真中に何が写っているかや、単体の物体や無生物が描かれた絵に対して何の絵かを識別する研究 [Eitz 2013] が進んでいる。しかしながら、これらの手法では描かれた絵の意味を捉えることはでき

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

ない。絵にモデルを付加する [Ito 2007] という観点ではアドベンチャーゲームなどである程度実現されているが学術的な体系化はなされていない。また、付加されている情報が人物偏重である、再利用される形で公開されていないという問題点がある。

2.2 本研究における絵の扱い

2.1 で述べた問題を解決するために、本研究では、絵モデルと呼ばれる絵に関する情報データを用いて、計算機による絵の扱いを実現する。絵モデルでは、絵内の各構成要素をオブジェクトと定義する。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、背景や吹き出しもオブジェクトとして扱う。

2.3 対象とする絵

すべての絵に対して合理的な絵モデルの定義は困難であるため、本研究では対象とする絵を限定して絵モデルを用いる。絵には動画と静止画があるが、今回は静止画の中で明確な時系列変化を持つ4コマ漫画を用いる。

2.3.1 4コマ漫画の利点

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る。人の知的生産物である4コマ内での変化をストーリーとオブジェクトの2方面から解析して統計的に見ることができれば役立つと考え、4章で具体的な提案手法を述べる。

2.3.2 自然言語の取り扱い

4コマ漫画には、必然的に自然言語による情報も付加されているが、本研究ではそれを積極的に用いない。ただし、アノテータは自然言語を含めて絵に情報付加をするため、ストーリー展開の分類などに影響がある点には注意を要する。絵情報と台詞等の自然言語情報の融合は重要な今後の課題である。

3. 従来手法

本章では、筆者らが提案中である4コマ漫画を対象とした絵モデルについて例を示し、問題点を明らかにすることで本研究の必要性について述べる。計算機において話の流れのある絵の構造化に焦点を当てるため、絵が主体となり話の流れを表現



図 1: 花火対決

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

```
<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>
<object>
<base id="3" class="花火" kind="線香花火" name="線香花
火" reality="true" coma="1, 2" />
</object>
... (中略) ...
<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)</emotion>
<action target="3">落とす</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"] を落とすたくない。 </think>
</character>
... (中略) ...
<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>
<icon owner="1">ショック</icon>
<icon owner="2">呆れる</icon>
<icon owner="1">滑る</icon>
<onomatopoeia owner="1">つるっ</onomatopoeia>
<event intentional="false">[link ref="5"] が落ちる</event>
... (中略) ...
</coma>
</story>
```

図 2: 作成した絵モデルの XML (一部抜粋)

している、話の流れが一般的に共感できる、という 2 つの観点を考慮して 1 つの 4 コマ漫画を具体例として選んだ。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 に絵の情報を表す XML を、表 1 に用いたタグの説明を示す。

表 1: 作成した絵モデルで用いたタグ (一部抜粋)

タグ名	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>	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。擬人化されている人間以外のものも含む。そ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>	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>	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>	表情。
<emotion>	感情。基本の 6 つの感情である、「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>	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。
<icon>	漫符の意味。
<onomatopoeia>	絵中のオノマトペ。
<text>	台詞。
<event>	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>	登場人物間の社会的・静的な関係性。

3.1 従来の情報付加法

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。内容は、物理的な情報と、社会的な情報、感情的関係などの非物理的な情報の大きく 2 つに分けて記述する。感情に関しては、ベクトルを用いて表す。

3.2 従来手法の問題点

まず、人手で付加する情報が多く、時間的なコストがかかるという問題がある。さらに、従来の絵モデルは粒度が細かいが、各オブジェクトの普遍性や状態変化の一般性に関する統計情報がないため、特殊な話の場合も、よくある話の場合も、どちらの情報も同等にしか扱えなかった。これは、各絵に対して一つのモデルを与えていたためで、例えば 4 コマ漫画の集合に対して頻出するオブジェクトや使われやすいストーリー展開のような点に着目していなかったためである。そこで、本研究では絵の集合に対して、ストーリー展開とオブジェクトに関する共通化部分を抜き出し、統計情報として再利用することを考える。この際にはオブジェクト自体の情報だけではなく、そのオブジェクトの時間的な変化についても考慮する。

4. 提案手法

本章では、提案手法について述べる。

4.1 基本方針

提案手法では絵を画像処理的な方法で理解するというアプローチではなく、ユーザと計算機の間で絵について自然な会話を成立させるために、事前に絵モデルを付加するというアプローチをとる。一般的に絵は実世界の写像であることが多く、本来絵に描かれているものをすべて理解するためには実世界に関する全知識が必要となる。しかしながら、一定のストーリーを有する絵の場合には、現れるオブジェクトの時系列変化やストーリー展開に一定のパターンが存在する場合が多い。そこで、提案手法では、これを具体的な情報としてデータベース化し、計算機が絵の内容を理解するための絵モデルに用いることを基本方針とする。

4.2 絵における時系列的な状態変化

提案手法では、絵の時系列的な状態変化を以下の2つとした。

オブジェクトの変化 ストーリーによらない、オブジェクト自体の変化。例えば、話に登場した食べ物が落ちる、食べられる等の変化である。オブジェクトに関するイベントも変化に含む。

ストーリーの変化 ストーリーに出現するオブジェクトによらない、話の展開における変化。ストーリー変化の共通化のため、オチに着目してストーリーの変化を分類する。

漫画には、「お約束」と呼ばれる定石化されたストーリー展開や誇張されたりデフォルメされたオブジェクトの扱いが数多く視られる。それを統計情報と持つことで絵を計算機で理解するための第一歩とすることを考えている。

本研究では、主にストーリーの変化に着目した。

4.3 オブジェクトの時系列的な状態変化

オブジェクトの情報は、変化しない情報である静的情報と変化する情報である動的情報の2つに分ける。

静的情報 静的な情報とは、一般的に時系列的に変化しない情報を指す。例えば、オブジェクトの辞書的な意味や、概念構造を表すための情報である。これらは既にインターネット上の辞書やコーパスとして、電子的な情報として提供されている。図1の具体的な静的情報としては、classの属性で示される「人」「花火」などが挙げられる。

動的情報 動的情報とは、フタが付いているものが開く、といったような物理的な変化と置かれている食べ物を誰かが横取りするなど、イベントの発生で捉えることができる変化である。

時系列的变化のある絵の各オブジェクトの状態変化と出現頻度を対としてデータベース化する。オブジェクトの変化としては、以下の二つが考えられるが、今回はオブジェクト自体が変化するパターンを調べた。

オブジェクト自体が変化するパターン <アイス>落ちる, <バス>進む etc.

オブジェクトに関するイベントを含めるパターン <ドア>[人が]頭をぶつける, <部屋>[人の]叫び声をする etc.

4.4 ストーリーの時系列的な状態変化

提案手法では話のオチからストーリーの時系列的な状態変化を以下の12種類に分類した。

- 0:失敗 登場人物が何かに失敗する話。
- 1:成功 登場人物が何かに成功する話。
- 2:勘違い 登場人物の考えが一般常識から外れていることを指摘されたり, 他者と考えが食い違う話。
- 3:無常観 時間経過や間を有して語られる悲しい話。
- 4:キャラクタ 登場人物の個性を強調する話。
- 5:大胆な行動 登場人物が大胆な行動を取る話。
- 6:意外な行動 登場人物が予想外の行動を取る話。
- 7:余計なお世話 登場人物の過剰な干渉により困る話。
- 8:学習 見聞きや経験により新たなことを学ぶ話。

9:文字 説明・セリフ依存。文字がないと内容が通じない話。

10:オチなし 「オチなし」とされる話。日常的变化の話, 登場人物の紹介, 商業的宣伝やナンセンスを含む。

11:その他 上記のどれにも当てはまらない話。

5. データの構築

本章ではデータ構築方針とその具体例を述べる。

5.1 情報の付加方法

オブジェクトのストーリー内での状態変化を調べるため、既存の4コマ漫画を選び、オブジェクトの状態をデータベース化した。今回選んだ4コマ漫画は「なめこタイムDX」(Web連載)と「スケッチブック1巻」から、各80話である。2名のアナテータが各作品を担当して情報を付加した。なお、4コマ漫画においては、話のオチが話全体の意味とも呼べる重要な情報であることから、オチの情報を付加した。話の構造としては、昔話の構造を31種に分類した手法[Propp 1987]が知られているが、現代の4コマ漫画には則さないため、4コマ漫画を数十話分析し、以下に示す12種のオチに分類することとした。

以下に情報付加の手順を示す。

5.2 情報付加の手順

以下に情報付加の手順を示す。図3に、以下の手順に沿って情報を付加した例を示す。

1. 1つの話の情報を以下のように記述する。

```
#ID:TITLE:Pi
<Oi>Kij:Sim
...
<On>Knj:Snj
%C
```

- ID:話数
- TITLE:話のタイトル
- P_i:オチの種類 [0:11]
- O_i:オブジェクト名
- K_i^j:コマ番号
- S_i^j:各コマでのオブジェクトの状態
- C:コメント

2. 次の話の情報との間に、一行の空行を挿入する。

```
#30:棒アイス:0
<人>1:[棒アイス]持つ,2:[ハサミ]持つ,3:[ハサミ]使う,4:[上部を開けて棒が持てないことに]がっかりする
<棒アイス>1:ある,3:[上部が]開く
<ハサミ>2:ある,3:[棒アイスの袋を]切る
... (中略) ...
#57:まどろみ:0
<人>1:まどろむ,2:眠る,4:こける
<バス>1:進む,3:止まる
```

図3: 情報付加方法

5.3 結果と考察

表2に各4コマ漫画から調べた話数, オブジェクト数, 人物を除いたオブジェクト数, オブジェクトの重複を除いた種数を示す。表3にオチの分類結果を, 表4にオブジェクトを6つのカテゴリに分けて分類した結果を示す。オチとして意味をな

表 2: 4 コマ漫画のオブジェクト情報の結果

	なめこタイムDX	スケッチブック
対象とした話数	80	80
オブジェクト数	169	162
オブジェクト数(人物除く)	45	56
オブジェクト種数	43	43
状態変化したオブジェクト種数	7	13

表 3: 4 コマ漫画のストーリー変化の分類の結果

オチの分類	なめこタイムDX	スケッチブック
失敗	8	10
成功	2	0
勘違い	4	4
無常観	14	0
キャラクタ	9	27
大胆な行動	4	3
意外な行動	3	3
余計なお世話	4	0
学習	0	0
文字	6	12
オチなし	19	16
その他	7	5

表 4: 4 コマ漫画のオブジェクトの分類の結果

	なめこタイムDX	スケッチブック
食べ物	10	5
生き物	1	2
機械	7	6
自然現象	4	2
無機物(機械以外)	11	23
未分類	10	5

すものとしては、失敗に関する話が多かった。2 作品の中で共通して出現したオブジェクトは「アイス」と「弁当」の 2 種のみであった。また、オチの分類において、その他に分類された 12 話のものは、アノテータが他の 11 種に分類しづらいと思った話だったが、改めて複数人で見直した結果、「アナロジー」と「パロディ」の分類を設けるべきであること、また、アノテータによる解釈の差を考慮する必要があることがわかった。

以下に今後検討を要する項目を記す。

粒度 オブジェクトをどの粒度で一般化するかについては、十分な検討が必要である。今回の例では、アイスというオブジェクトが話によって「食べ物」、「袋に包まれているもの」、「くじ付きのもの」など複数の特徴で分類すべき場合があった。これは話の展開によってオブジェクトを見る粒度が変化することを示唆している。今回の結果から、表 4 のように分類したが、データベースの大規模化に伴って、適切な一般性を保ちながら細分化されすぎないクラスタリング手法を考える必要がある。

部分と全体の関係 オブジェクト付いている「フタ」を開ける場合、「フタ」はオブジェクトの一部である。また、食べ物を半分食べる話では、半分になったことは重要な情報だが、オブジェクト名に変化はない。

このような部分と全体の関係や has-a 関係の情報をどのように記述するかは難しい問題である。

前提知識 話を理解するために必要な前提知識が多い。例えば、棒アイスの袋の反対側を開けたことにより、持ち手である棒を持つことができずがっかりする話では、「アイスの片側に棒がついている」、「アイスは冷たい」などの前提知識が必要である。また、人物の行動が意外なものかどうかは、一般的な行動のどれにも合致しないか、人がそのような行動を取ることが奇異だということを前提知識

として知っている必要がある。例えば、「キャンパスの前で念を送る」という行動が奇異かどうかを判断するには、「キャンパスの前」という状況で人が一般的に何をするか(例:キャンパスに絵を描く、キャンパスに描かれた絵を見る)を知っているか、「念を送る」という行動が奇異な行動の典型例と知っている必要がある。

異なる解釈 数名の被験者に 4 コマ漫画を見せた結果、話の解釈が大きく異なるものがあった。漫画をよく読むかや、育ってきた環境でそのような経験があるか、文化的背景などに依存して解釈が変わる話がある。また、年齢によっても理解できるか否かが異なる。このような解釈の異なりがどのオブジェクトに由来するかということが、絵を見て話を理解する上での重要な手がかりになると考えるため、今後調査を進める予定である。

感情依存 驚いた後哀しい気持ちになる、大胆な行動を取りたい衝動に駆られた後冷静になるのように、感情の変化による話のパターンが多く見られたため、感情の変化を抽出し、データベース化する手法が必須である。

キャラクタ設定の反映 スケッチブックの作品では、特にキャラクタの関係性に依存したオチになることが多かった。このためキャラクタ設定を適切に表現できれば、性格に沿った行動である、性格に反する行動である、といった情報が手がかりとなる可能性がある。

6. まとめと今後の課題

本研究では、工学的に絵の情報付加をするために、従来の絵のモデル化手法の問題点を示した上で、絵の時系列的な変化に着目してデータベース化する提案手法を述べた。情報付加する中で共通要素を検討し、記述方法を整備することを今後の課題とし、時系列的変化から絵の意味を理解することを目指す。また、応用として、本研究で構築したデータベースを参照し、ユーザ入力をもとに事前に用意したパーツを自動的に組合せて漫画を作成するシステムを発表時に示す。

参考文献

- [Ueno 2012] Miki Ueno, Naoki Mori, Keinosuke Matsumoto : *Picture Information Shared Conversation Agent: Pictgent*, Distributed Computing and Artificial Intelligence Advances in Intelligent and Soft Computing, Springer, Volume 151, pp.91-pp.94, (2012)
- [新井 2012] 新井 紀子, 松崎 拓也 : ロボットは東大に入れるか? 一国立情報学研究所「人工頭脳」プロジェクト一, 人工知能学会誌, Vol.27, No.5, pp.463-pp.469(2012)
- [Eitz 2013] Mathias Eitz, James Hays and Marc Alexa : *How Do Humans Sketch Objects?*, ACM Trans. Graph. (Proc. SIGGRAPH), vol.31, pp.44:1-pp.44:10, (2012)
- [Ito 2007] Kazunari Ito, Motohiro Matsuda, Martin J. Durst and Koiti Hasida : *SVG Pictograms with Natural Language Based and Semantic Information*, SVG Open (2007)
- [Propp 1987] Vladimir IAKovlevich Propp (著), 北岡 誠司 (訳) : 昔話の形態学, 書肆風の薔薇 (1987)