

時系列現象を表現する絵とストーリー間の定量的解析

Computer Analysis between Sequential Pictures and Stories

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, Osaka Prefecture University

Creative sequential pictures such as comics and picture books contain a lot of useful representation of pictures to depict stories. However, it is difficult for computer to understand these deformed pictures based on computer vision techniques, therefore few analysis between representation of pictures and stories have been reported in this field. Although we have proposed models of pictures for computers, lots of difficult tasks to modeling suitable pictures are have been still remained. In this study, we focused on state of characters in 4-scene comics, in which pictures are divided into fundamental parts. We propose the novel features of pictures by utilizing direction of character faces and types of balloons.

1. はじめに

計算機による人の知的な創作物の理解, 生成は, 人工知能分野の大きな目的の1つである. 創作物には, 小説, 絵画, 音楽など様々な媒体があり, 人が意味づけた文字や線, 図形, 音符など異なる記法で構成されている. 局所的に見ると, それらはすべて点の集合であるが, 人は意味的なまとまりである, 文字や図形, 記号などの種類に分けて理解することができる. しかし, 人がどのように意味的に理解しているかは明らかになっておらず, 中でも自由度が高いストーリーを有した創作物を対象として人の理解を定量的に示すことは非常に困難 [1] である. 著者らはこれまで, 文字ではなく, 国籍や年齢に寄らない表現手段である絵に着目しており, 絵を用いたストーリーを計算機に理解させる, また, 自動生成することを目指している. 絵を用いたストーリー中で, 4コマ漫画はあらゆるストーリーを作成することができるという創作の自由度は保っているが, コマの長さが一定であり, 作者が意図的に強調したいものを取捨選択してデフォルトするため, ストーリーが比較的明確であり, 工学の対象 [2] にしやすいと考えられる. そのため, 4コマ漫画を対象として, 解析と生成の両面から研究を続けている. これまで, 漫画に出現する無生物オブジェクトとストーリーパターンの関係性の解析 [3] や, 状態変化をモデル化する方法を提案 [4] してきた. しかし, 無生物を対象とした研究では, 4コマ中の無生物の種類別と状態をどの粒度で記述するかという問題や, 背景としての無生物など, 直接的にストーリーに関わる無生物が登場しないという問題があった. 外的な力によってしか変化しない無生物のみでストーリーが進行することは少ない. 本研究では, 4コマ漫画のストーリーのパターンに強く関連する絵の特徴の時系列データを定量化する方法を提案し, 考察した. 絵の特徴として, 今回は登場人物の顔の向きと吹き出しを対象とする.

本研究の構成は以下の通りである. 2. 章に関連研究を示し, 3. 章で提案手法を述べる. その後, 4. 章で具体的なデータを作成して実験し, 5. 章で今後の課題を述べて, まとめる.

2. 関連研究

ストーリーの定量化と絵の定量化の2点について, 関連研究とその問題点を以下に示す.

2.1 ストーリーの定量化

ストーリーを定量化するためには, ストーリーの部分や全体を何種類かに分類する必要がある. ロシアの昔話の整理方法として 31 種 [5] の分類が知られる. しかし, この方法は 4コマ漫画のストーリーの分類には適さない. 筆者らは 4コマ漫画のオチの分類を試みたが, すべての 4コマ漫画を過不足なく分類することは非常に困難である.

2.2 絵の定量化

絵の定量化には, 研究目的によって, いくつかの方法が存在する. 例えば, 絵を画像認識してオブジェクト種別を調べる手法 [6], 絵画や動画の検索に有用なタグを付与する手法などである. 最大の問題は, タグ付けの粒度と記法の定義である. 例えば, 「ネコ」とタグが付与されていても, 具体的にネコが絵中でどのような状態なのかは分からない.

筆者らはこれまで, 絵中の要素すべてをオブジェクトとみなして 1 枚の絵を意味的なまとまりに分けようと考えた. 4コマ漫画中のオブジェクトとは, 登場人物, 吹き出し, 漫符, 台詞, 背景などである. しかし, 以下のような課題が見られた.

- 無生物についてオブジェクト種別と状態変化を分類し頻度を調べたが, 無生物が出現するストーリーの数は少なく, 特徴を発見するだけの十分なデータ収集が困難である.
- データ作成にアノテータの主観が入る.
- オブジェクトの部分と全体を区別することに課題がある. 部分と全体の区別の課題の例を挙げると, まず背景が挙げられる. 人は, 机, 椅子, 黒板などの複数のオブジェクトの組合せでできているものを, ときに各オブジェクトの単体とみなし, 時に, 背景という全体とみなす. また, アイスとアイスの棒や, ペットボトルとペットボトルのフタなども, 部分と全体の区分が困難な例である.
- 枠内のスペースは限られているため, 登場人物の全身など, 特定のオブジェクトの全体をコマの枠内に収めるこ

とはあまりなく、登場人物の顔や手など特定の部位だけを枠内に描くことが多い。そのため、計算機で画像認識的にデータを自動構築することは困難である。一方で、人がすべてアノテーションするのは現実的ではない。

そこで、これらの課題に依らない新たな特徴を発見する必要がある。

3. 提案手法

関連研究で述べた問題点を解決するため、本研究では新たに、絵表現に着目したストーリーのパターン化のため、登場人物の顔の向きと吹き出しの形状という4コマ漫画の特徴量を提案する。

3.1 絵の特徴の候補

まず、4コマ漫画のストーリーのパターンを表すための絵の特徴の候補を以下に分類し、紹介する。

オブジェクト種別 漫画の中でのオブジェクトの種別の分類。登場人物、無生物、吹き出し、セリフ、漫符、効果線、描き文字など。各上位クラスをどう定義するか、どこまで詳細に下位クラスを定義するかは困難な課題である。体系化が十分になされておらず、膨大な種別があり、意味的に分類することが困難である。漫画に特有でない情報に関しては、Wordnet[7]など既存のオントロジーを参考にすることができるが、以下の部分と全体の問題がある。

オブジェクトの状態 人物であれば、顔や体の向き、立っている、座っているなどの姿勢、笑っている、泣いているなどの表情など、無生物であれば、割れている、溶けているなどの状態が考えられる。状態の種別だけでなく、激しく泣いている、など程度をどこまで細分化するかは難しい問題である。

漫画特有のオブジェクトの種別を項目立て、それぞれに対し考え得る下位クラスの具体例を示す。

1. 漫符 ひらめき、汗、怒り、照れ、涙、青筋、震えなど、主に人物の精神状態を表すために使用される。
2. 擬音語 静けさを表す「しーん」や、新たな人物の出現や怒気を表す「ゴゴゴゴ」など。主に雰囲気や人物の感情などを表すために使用される。
3. 効果線 特定オブジェクトの強調やオブジェクトの動作の軌跡を表すために使用される。
4. 吹き出し 通常台詞、大声、小声、独白、空想など、吹き出しの形状で、声の大きさや吹き出し内の文字が何を表すものかを示す。

3.2 基本方針

上記で、モデルを作成するにあたって、漫画の絵の特徴となり得るものを挙げた。すべての絵の特徴を定量化して、ストーリーとの相関を調べることは理想的だが、絵中のオブジェクトの種類や状態の分類は難しい。そこで、本研究では、以下の3点の考えに基づき、4コマ漫画の絵の特徴量の中で、登場人物の顔の向きを系列データとし、ストーリーを代替する方法を提案する。

- 迅速に大規模に工学の分野に役立てるため、データの作成に人の主観や複雑な思考過程が介在せず、画像認識の比較的近い進歩で自動データ化が可能となりそうな特徴が望ましい。現在は完全な自動化が困難であるが、人手でもアノテータが迷わず付与できる、簡素に抽出でき、かつ情報量が多い特徴が求められる。

- ストーリーには何らかの時系列的な変化がなければならぬため、4コマ漫画のストーリーの中で中心的な役割を果たすオブジェクトは自らの意思を持って行動ができる登場人物であることが多い。

- 4コマ漫画は各話のコマ数が一定のため、文字数が決められた俳句や短歌のように、ストーリーのテンポが重要である。画面内で比較的広い領域を占める、登場人物、またその状態変化の代表的なものとして、登場人物の顔の向きがストーリーの意味や印象に影響する可能性が高い。

さらに、以下の3点の考えに基づき、吹き出しの形状に関する要素もデータ付与した。

- 漫画の中心的な要素は絵だが、台詞も漫画中で重要な意味を占める。
- 吹き出しの領域識別 [8] の研究が知られており、吹き出しの形状は自動取得できるデータとなる可能性が高い。
- 台詞など吹き出しの内容の意味理解は自然言語処理の分野でも困難とされているが、吹き出しの形状だけで吹き出し内の文字の種別は分類できる可能性があり、形状は重要な特徴である。

3.3 データの作成：絵の定量化

以下に、データの作成方法を示す。

データ作成方法

$$\#n : t : o_1 > d_1^{o_1} d_2^{o_1} d_3^{o_1} d_4^{o_1}, o_2 > d_1^{o_2} d_2^{o_2} d_3^{o_2} d_4^{o_2} : c_1^1 c_2^1 c_3^1 c_4^1 c_5^1, c_1^2 c_2^2 c_3^2 c_4^2 c_5^2, c_1^3 c_2^3 c_3^3 c_4^3 c_5^3, c_1^4 c_2^4 c_3^4 c_4^4 c_5^4$$

n 作品番号。

t 各話のタイトル。

o_j 登場人物番号 j のオブジェクト名。

$d_i^{o_j}$ i コマ目の登場人物オブジェクト o_j の顔の向き。

c_k^i i コマ目中の吹きだし種別番号 k のオブジェクトの出現の有無。1, 0 の2進とする。

登場人物のオブジェクト種別 o_j の記法 同一のオブジェクトが出現する場合は、オブジェクト種別の後ろに番号を伏して区別する。例えば、今回は登場人物だけを対象としているので、2人の人物が登場するときは、 o_1 に「人1」、 o_2 に「人2」と記述する。出現位置が先のものから昇順に番号を振る。出現位置は、4コマ漫画全体の中ではコマ番号が小さい方が先、同一コマの中では、登場人物領域の右端が右側のものを先とする。これは、対象とする日本の4コマ漫画の読む順序を考慮しており、コマを上から下へ、コマ内では右から左に読むことを想定しているためである。

登場人物の顔の向き $d_i^{o_j}$ の記法 登場人物 o_j の顔の向きを画面前から見て、以下の8種類のいずれか1つに近似的に割り付ける。

- 0 該当コマにおいて人物の出現なし
- f 前, b 後, r 右, l 左, u 上, g 下
- p 該当コマにおいて顔の出現なし

吹き出しの種別 c_k^i の記法 以下の5種に近似的に分類する。吹き出し種別番号と形状特徴の対応を示し、図1に吹き出し種別を示す。種別番号は k に該当する。

- 1. 楕円形枠
- 2. ジグザグ系枠
- 3. 四角形枠
- 4. その他枠形状
- 5. 吹き出し枠で囲われない活字

1~4の漫画中の活字を囲う吹き出し枠と5の活字を対象とする。文字認識の技術で取得が容易な文字が望ましいため、作者が手書きした装飾的な文字は今回対象外とした。

3.4 例外処理

以下、データ作成に注意を要する点を説明する。

登場人物の定義 創作物では人間だけではなく、動物や擬人化された無生物が出現することも多いため、人のように主体を持って振舞うものを登場人物と呼ぶ。

特定のコマに登場人物が出現しない場合 顔も出現しないので0に割り当てる。

顔が出現しない場合 コマ内に手や足など特定の部位だけが描かれることがある。顔の向きはわからないため、pに割り当てる。いずれの人物の部位かは前後の文脈から判断できるが、コマ単位での判別は難しい。そのため、人は情報を付すことができるが、現在の情報処理技術での自動付与は難しい。

顔と体の向きが異なる場合 顔の向きだけを対象とする。

2方向以上を含み判別に迷う場合 最も近い方向に近似する。

3.5 具体例

図2で本研究の説明用の4コマ漫画を例にとり、図3にデータ付与の具体例を示す。最終的には画像認識的手法を用いることにより、データ作成の自動化を目指す。今回は、人手でアノテーションした。

4. 実験

提案手法の有効性を調べるため、以下の手順で既存の漫画を対象として、データセットを構築して頻度を調べ、考察した。

- データを付与する対象の作品として、スケッチブック1[9]の冒頭より80話を使用した。
- 3名の本大学の情報系の大学院生に記法を説明し、データを作成させた。
- 4コマ漫画各話の登場人物が2人以下のものを対象とし、3人以上のものは省いた。そのため、 o_1 が登場人物1、 o_2 登場人物2を表す。以下の結果では、向きは80話中66話のデータを示す。

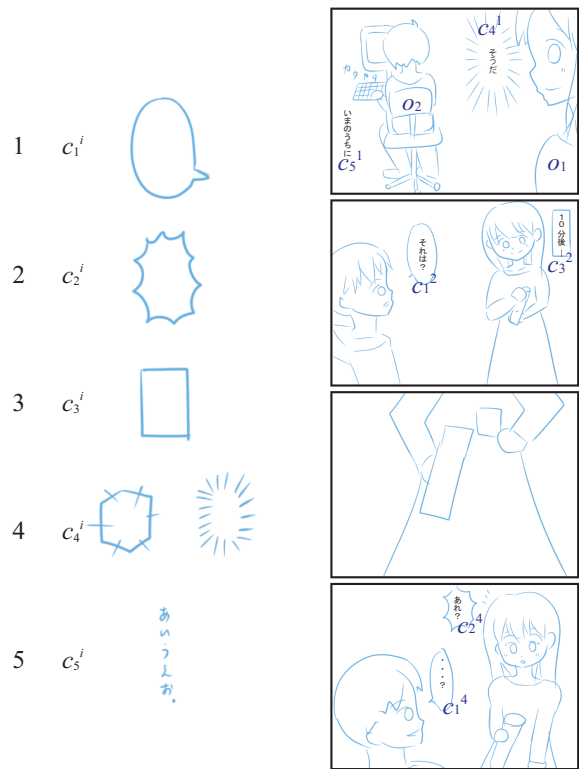


図1: 吹き出し種別の例

図2: 登場人物の向きと吹き出し形状付与例の4コマ漫画

#1:作戦:人 1>lfpf, 人 2>br0r:00011,10100,00000,11000

図3: データ付与の具体例

表1: 80話中の登場人物数ごとの話数
表2: 対象とした66話中の o_1 と o_2 の各向き $d_i^{o_j}$ の頻度

登場人物数	話数
0	0
1	20
2	46
3	10
4	2
5	0
6	2
計	80

向き	$d_1^{o_1}$ の頻度 (順位)	$d_1^{o_2}$ の頻度 (順位)
0	26 (4)	46 (2)
f	138 (1)	63 (1)
b	20 (5)	10 (5)
r	38 (2)	36 (3)
l	38 (2)	26 (4)
u	0 (8)	0 (8)
g	2 (6)	2 (6)
p	2 (6)	1 (7)

3名中1名のアノテーションの結果を具体例として示す。表1に4コマへの登場人物数別の話数を示す。表2に o_1 と o_2 の向きの頻度と順位を示す。表3に対象とした66話中の o_1 と o_2 の系列データの種別と頻度を示す。表4に各コマの吹き出し系列種別と頻度を示す。

- 各登場人物の向きのデータ系列 $d_1^{o_1} d_2^{o_1} d_3^{o_1} d_4^{o_1}$ が取り得る場合は $8^4 = 4096$ 通りあるが、今回データ付与した全112登場人物分の向きのデータ系列は67通りであった。これは全組合せの1.6%にあたる。今回、取得したデータが少ないため、同作品の他のデータや、他作品のデータを十分に取得し、さらに精査にする必要がある。大規模化してなお疎な構造であれば、漫画のストーリーを向きでパターン化することが非常に有効な可能性が高い。
- 吹き出しの向きのデータ系列 $c_k^1 c_k^2 c_k^3 c_k^4$ が取り得る場合は $2^0 = 1048576$ 通りあるが、今回データ付与した、全80話中に出現したデータ系列は61通りであった。さらに、

表 3: 対象とした 66 話中の各 o_i の向き系列 $d_1^{o_i} d_2^{o_i} d_3^{o_i} d_4^{o_i}$ 種別と頻度

o_1		o_2		$o_1 \cup o_2$	
頻度	系列	頻度	系列	頻度	系列
10	fff	4	fff	11	fff
3	frr	3	rrr	5	fff
3	ffo	2	rff	5	fff
3	fof	2	ffr	4	fff
2	lll	2	ffo	3	fff
2	llf	2	ofr	3	fff
2	frl	2	oof	3	fff
2	frf	2	00ob	3	fff
2	flr			2	fff
2	flf			2	fff
2	ffl			2	fff
2	ffbb			2	fff
2	fbbb			2	fff
2	f000			2	fff
(頻度 1 以下の系列は記載省略)					
42 種		35 種		67 種	

表 4: 対象とした 80 話中の i コマ目の吹き出し系列 $c_1^i c_2^i c_3^i c_4^i c_5^i$ 種別と頻度

c_k^1		c_k^2		c_k^3		c_k^4		$\sum_{i=1}^4 c_k^i$		$c_k^1 c_k^2 c_k^3 c_k^4$	
頻度	系列	頻度	系列	頻度	系列	頻度	系列	頻度	系列	頻度	系列
27	00000	28	00000	28	00000	23	10000	104	10000	7	10000,10000,10000,10000
27	10000	26	10000	28	10000	17	00000	100	00000	5	00001,00001,00001,00001
11	00001	9	00001	6	00001	9	00001	35	00001	5	00000,00000,00000,00000
6	00010	6	00010	4	00100	8	10010	21	00010	3	00000,00000,10000,10000
4	01000	5	10010	3	01000	7	00010	18	10010	2	10000,10000,10000,00000
2	10010	2	00100	3	10010	4	01000	12	01000	2	10000,00000,00000,00000
1	00100	1	01000	2	00010	3	00100	10	00100	2	00000,00000,00000,00100
1	10001	1	01001	1	00101	2	00101	5	11000		(頻度 1 以下は系列記載省略)
1	11000	1	10001	1	00110	2	01001	4	01001		
				1	01001	2	01010	3	00101		
				1	01100	2	11000	3	10001		
				1	10001	1	10100	2	01010		
				1	11000			1	00110		
								1	01100		
								1	10100		
9 種		10 種		13 種		12 種		15 種		61 種	

頻度 2 以上のものは、7 通りのみであった。これは全体の約 $6.68 \times 10^{-4}\%$ (小数点第 2 位以下四捨五入) で極めて疎な構造であり、ストーリー特徴を表すパターンとして有効な可能性が高い。大規模化して調査を進める。

- 表 2 から、 o_1 も o_2 も f が 1 位であり、正面の顔が多いことがわかる。今回、額と髪など、顔の一部分でも含まれていれば向きを付したが、多くの場合、目が含まれていたため、画像認識で顔抽出がしやすい画像と言える。また、正面の場合、目、鼻、口の表情が明確に分かるため、今後表情を特徴とすることができると考える。
- o_1 と o_2 に共通で最も頻度が高い fff の系列は主人公の動きが乏しく、文字の台詞主体で話が進む傾向が見られた。
- o_1 と o_2 の向きの頻度の差異に着目すると、 o_2 はコマへの出現のないことを表す 0 の頻度が高いことから、 o_1 が話の主人公であり、 o_2 が補佐的な役割を果たしていると仮定し、向きの系列から主人公を予測できる可能性がある。

5. まとめと今後の課題

本研究では、各コマの登場人物の向きと吹き出しの形状という絵の特徴に着目して計算機に親和性の高い 4 コマ漫画のストーリーを代替するデータを提案した。提案手法に基づき、既存の漫画を対象にデータを作成して話や他の絵の特徴との関係を考察した。本稿に向きや吹き出し形状の頻度を示したが、他に得られた詳細な結果やアノテーション間の差異については発表時に示す。今後の課題を以下に挙げる。

- 向きと吹き出しの特徴が付しやすいことが明らかになった。他作品についても解析を進め、大規模データ化し、精緻な定義をするまた他オブジェクトについても調べる。
- 向きを角度情報として表し、吹き出し形状の種別など他の絵の特徴と合わせ素性ベクトルとして機械学習することで、漫画全体のオチや作品名、作者名などの未知クラスを学習する方法を提案する。
- 4 コマ漫画の作者にストーリーを作成するときに重視する項目を問い、作者の中で体系化された特徴を調べる。

謝辞

本研究は JSPS 特別研究員奨励費 (課題番号 25・10978) の助成を受けたものです。

参考文献

- [1] きまぐれ人工知能プロジェクト 作家ですよ。 http://www.fun.ac.jp/~kimagure_ai/.
- [2] 松下光範. コミック工学の可能性. 第 2 回 Web インテリジェンスとインタラクション研究会, <http://amateras.wsd.kutc.kansai-u.ac.jp/wordpress/wp-content/uploads/2013/05/WI22nd-mat-63-68.pdf>, 2013.
- [3] 上野未貴, 森直樹, 松本啓之亮. 絵の時系列的状態変化分布に基づく絵モデルの構築. 人工知能学会, 203-13 in, 2013.
- [4] Miki Ueno, Naoki Mori, and Keinosuke Matsumoto. 2-scene comic creating system based on the distribution of picture state transition. In *Distributed Computing and Artificial Intelligence, 11th International Conference*, Vol. 290 of *Advances in Intelligent Systems and Computing*, pp. 459-467. Springer International Publishing, 2014.
- [5] Vladimir Iakovlevich Propp. *Morphology of the folktale*. University of Texas Press, 1968.
- [6] Mathias Eitz, James Hays, and Marc Alexa. How do humans sketch objects? *ACM Trans. Graph. (Proc. SIGGRAPH)*, Vol. 31, No. 4, pp. 44:1-44:10, 2012.
- [7] George A. Miller. Wordnet: A lexical database for english. *Commun. ACM*, Vol. 38, pp. 39-41, 1995.
- [8] 田中孝昌, 外山史, 宮道壽一, 東海林健二. マンガ画像の吹き出し検出と分類. *The Journal of The Institute of Image Information and Television Engineers*, Vol. 64, No. 12, pp. 1933-1939, 2010.
- [9] 小箱とたん. スケッチブック 1. マッグガーデン, 2003.