

## 漫画内の特徴的要素が与えるストーリーの印象についての検討

## Relationships between Features and Story Description in Comics

上野 未貴\*<sup>1</sup> 井佐原 均\*<sup>1</sup>  
Miki Ueno Hitoshi Isahara

\*<sup>1</sup>豊橋技術科学大学, 情報メディア基盤センター  
Information and Media Center

Automatic analysis and creation of intellectual products such as comics is one of the important topic in the field of artificial intelligence. However, it is difficult to understand comics by basic computer vision techniques because they have emotional stories drawn by complex representation. In this study, features of comics are classified and discussed what features are important. To confirm the effectiveness of proposed method, computer simulations are carried out taking data of four-scene comics as examples.

## 1. はじめに

表現学習の飛躍的な進歩により, 人工知能分野で計算機による漫画という創作物の理解や生成への期待 [1] が高まっている. 創作物の中でも, 連続した時系列的变化によるストーリーを有する創作物, 小説や漫画に関する研究が増加している. しかし, このような創作物の理解生成には, 段階を踏んだ理解生成, 状況理解, 感情理解, 連続した状況予測などが必要と考えられるが, 必須な要素技術やその組合せがわかっておらず, 工学的に問題を設計すること自体が非常に困難である. 本研究では, 絵と文字の複合的な媒体であり, ストーリーを有した創作物である, 漫画に着目する. 我々が漫画を読む際にストーリーを理解するために用いる漫画内の特徴, 知識は明らかでない. そもそも, 何を以て理解したと定義するかが難しい. 本研究における漫画の理解とは, 「漫画を見て, ストーリーを単なる事象だけではなく感性的な情報を含めて説明すること」と定義する. すなわち, 漫画のデータを入力として与えて, 適切なストーリー印象クラスに分類したり, 適切な説明文を出力することに相当する. そこで, 本研究では, この入力画像と出力クラスを定義し, その関連および過程を, 人による実験および計算機的な数値実験により検討することを目的とする.

本研究の構成を以下に示す. 2. に本研究で対象とする特徴の基本方針を, 3. に提案手法を示し, ストーリークラスの定義方法を述べる. その後, 4. で実験について詳細に説明し, 5. で Deep Learning を用いた実験について示し, 6. で今後の課題を述べまとめる.

## 2. 基本方針

人の漫画の理解過程を考え, 関連する研究と対応付けながら述べる. まず, 漫画という表現媒体によらない絵の表現について考える. 例えば, 二者が向かい合って手を差し出し握りあっていたら, 「握手している」と理解できるように, 日常生活で見る行動がそのまま絵として表現されている場合である. このような日常生活で見ることでできる表現の場合, 大量の写真データから学習して説明文を出力する画像理解の研究は近年の Deep Learning [2] の適用により飛躍的に進歩している. また,

連絡先: 上野 未貴, 豊橋技術科学大学, 情報メディア基盤センター, 〒441-8580 愛知県豊橋市 天伯町雲雀ヶ丘 1-1, E-mail: ueno@imc.tut.ac.jp

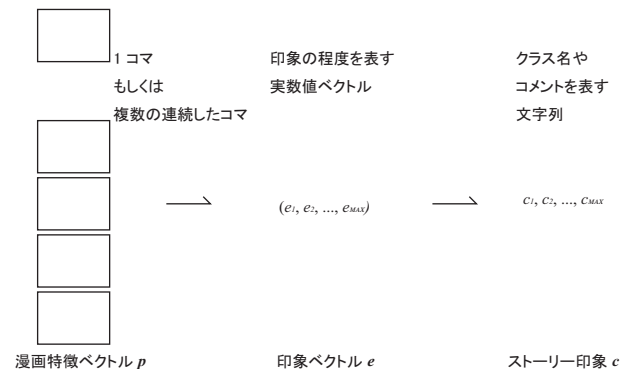


図 1: 研究目的

写真中の人物の表情を識別して感情値を出力する研究も進んでいる. 一方で, 漫画特有の表現もある. 例えば, 額の横あたりに縦線が複数本入ってれば落ち込んでいる, 星のようなマークに囲まれた対象物は, 綺麗な, あるいは立派なものである, というようなことを表す. しかしながら, 漫画の絵の表現形態を知らない者には理解できない. また, 画像中に吹き出し [3] があり, 台詞が併記されることも漫画の特徴である. また, 漫画の理解には, その事象が起こりやすいことか否かなど, これまでの人工知能分野の一般知識獲得の問題も内在している. すなわち, 漫画の理解に必要な特徴は以下の 3 つである.

1. 一般的な画像特徴
2. 漫画固有の画像特徴
3. 背景知識

本研究では, 特に, 漫画の画像内に明示的に現れている特徴を対象とし, 一般知識獲得の問題は副次的に考察する.

## 3. 提案手法

### 3.1 概要

漫画の理解を工学的な過程として表すと, 漫画の 1 コマや複数の連続したコマ  $q$  を入力として, 受けた印象  $e$  を文字列  $c$  として出力することとする. 図 1 に本研究の目的の概略図を示し, 達成すべき課題を, 図と対応付けて以下に示す.

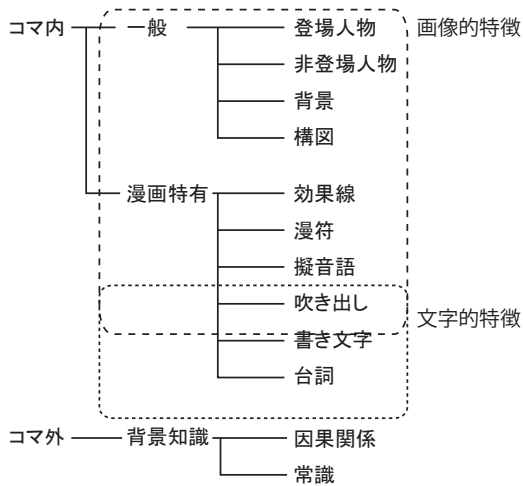


図 2: 漫画の特徴的要素

1. 漫画  $q$  の特徴のベクトル表現  $q$  の定義
2.  $\xi: q \rightarrow e$   
漫画  $q$  の特徴からストーリー印象  $e$  への写像の定義
3.  $\psi: e \rightarrow c$   
ストーリー印象  $e$  からストーリークラス  $c$  への写像の定義

本研究では、 $q$  と  $c$  を詳細に定義するための方法を考察する。 $e$  についても具体的なベクトル表現で表して、写像  $\psi$  について考慮すべきだが、今回は簡略化のため  $\psi$  を単射とする。

本研究では、コマの長さが決められており、明瞭なストーリーが多いことから、工学的に扱う対象として好ましいと判断し、4コマ漫画を対象として漫画理解について検討する。4コマ漫画の場合、4コマ目がストーリーの印象全体を理解する上で重要な、いわゆるオチのコマを決定づけることが多い。そのため、 $q$  を4コマ漫画の4コマ目とし、4コマ目の印象と画像特徴の印象の関係を調べる。台詞特徴の影響を考慮するため、台詞に関する特徴的要素の数を変更した3種のデータセットを用意する。人の  $q$  と  $e$  の対応付け、また画像の表現特徴の獲得に強い畳込みニューラルネットワークを用いて比較実験をし、画像特徴以外の台詞や背景知識、人物間の関係性などの情報がどの程度必要かについて考察する。

### 3.2 漫画の特徴的要素

漫画という媒体を表現するために必要な特徴要素集合を仮定する。漫画を理解する上で必要な漫画中の特徴要素が過不足なく揃う集合を工学的に表す必要がある。漫画では作者が意図的に描くものを選択しているため、写真よりもストーリー理解に無関係な要素は少ないが、修飾的な要素も含まれている。

特徴ベクトル  $q$  を定義する前に、漫画の特徴的要素を区別して示す。図2に、人が一般的に漫画を理解する上で重要と考えられる特徴的要素を、コマ内に表れている情報と、コマ外の情報に分けて示す。大きく分けて、コマ内の画像に依存した特徴、コマ内の文字に依存した特徴、コマ外の常識、背景知識の3つである。さらにコマ間の特徴変化によって、ストーリーを理解する必要がある。

### 3.3 漫画の特徴ベクトル $q$ の定義

漫画の特徴ベクトル  $q$  を定義するため、漫画中のオブジェクトを一部欠損させることにより、いずれの特徴が重要か考察する。画像全体の画素濃度を画像特徴ベクトルの要素とする場

合、特定オブジェクトの領域に該当する要素を0とすることに相当する。欠損させるオブジェクトは3.2で挙げた中で複数の組合せが考えられるが、近年の画像認識技術の発展により、画像特徴のみに依存する特徴については自動獲得できる表現が多いと考えられるので、画像特徴と他の情報を区別して扱う。背景知識を意図的に欠損させることは懇談であり、漫画は、画像と文字のマルチモーダルな媒体であることが特徴であり、特に4コマ漫画では画面のサイズが制限されていることから、文字が有する情報も多い。そのため、まず、画像特徴と台詞特徴を分離して考察するため、台詞特徴を段階的に欠損させ、下記の3種のデータセットを構築した。図3に各データセットの画像の例を示す。

1. オリジナル画像
2. 台詞なし画像
3. 吹き出し統一画像

吹き出し統一画像作成中に、吹き出し内にハートのマーク、疑問符など、感情をアイコンで表した場合や、想像している内容の絵が描かれていたり、吹き出しの中身が必ずしも文字とは限らない場合があった。

### 3.4 コマクラス $c$ の作成

人の漫画理解過程に近づけるため、必要なクラスを考え、正解データの作成方法を検討する。各データセットの画像を与えて、以下の指示でアノテータにクラスを作成させる。

1. 画像にシーンの印象を表す10文字以内の文字列を付与する。本研究では、この文字列をタグと呼ぶ。なお、タグは複数付与して良い。
2. すべての画像にタグを付与した後、全体を見直してなるべく複数画像間で共通しており、なおかつその画像をよく表すと考えられるタグを各画像に1つのみ残し、クラスとする。

## 4. 実験1: 人の漫画識別

### 4.1 データセット構築

まず、本研究ではシーン内の画像特徴を十分に考慮するため、データセット構築に、ある既存漫画の1巻に収録された話の4コマ目のみを用いた。画像枚数は、188枚である。画像サイズは290×210ピクセルである。

### 4.2 実験設定

3種のデータセットに対し、アノテータ(20代男性)に各2回タグを付与させた。所要時間は、1種のデータセットあたり約1時間程度であった。なお、記憶しないよう、日を空けて、また、漫画の特徴的要素の少ない順、すなわち、吹き出し形状統一→台詞なし画像→オリジナル画像、の順に実施した。以降、以下の変数を用いて説明する。

- $s_i^j$ :  $j$  番目のデータセット中の  $i$  番目の画像。データセット番号  $j = 1, 2, 3$  は3.3で示した1.~3.の番号を示す。
- $N$ : 各データセットに含まれる画像枚数。
- $C(s_i^j, t)$ :  $t$  回目に  $s_i^j$  の画像に被験者が割り当てたクラス
- $M(m_1, m_2)$ :  $m_1$  と  $m_2$  が一致するとき1、異なるとき0を返す関数。

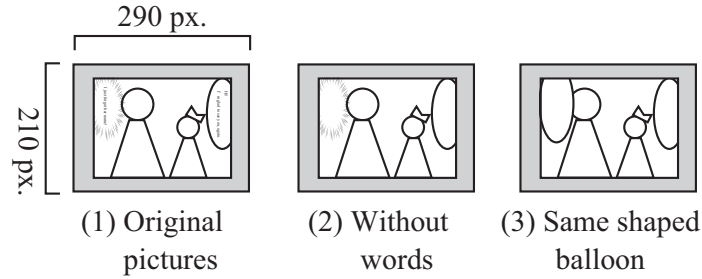


図 3: 3 種のデータセットの画像例

表 1: 各データセットの試行間一致率

オリジナル画像	台詞なし画像	吹き出し統一画像
0.66 (= 124 / 188)	0.59 (= 111 / 188)	0.47 (= 89 / 188)

表 2: 各データセット間の一致率

$t$	$a = 1, b = 2$	$a = 1, b = 3$	$a = 2, b = 3$
1	0.52 (= 98 / 188)	0.45 (= 84 / 188)	0.48 (= 90 / 188)
2	0.48 (= 91 / 188)	0.37 (= 70 / 188)	0.43 (= 81 / 188)

表 4: 一致種別別頻度

一致種別	頻度	
全一致	28	
部分一致	$a = 1, b = 1$	8
	$a = 1, b = 3$	2
	$a = 2, b = 3$	3
不一致	2	
計	43	

表 5: 全一致のクラスと頻度

クラス	頻度	特徴
怒り	8	登場人物複数.. 吊り肩, 青筋, 怒りを表す漫符
呆れ	8	登場人物複数, 一人笑顔, 一人呆れた表情.
ほのぼの	5	登場人物複数, 全員笑顔.
驚き	4	顔周りの驚きを表すジグザグの漫符.
困惑	3	焦りを表す汗, 顔周りの疑問の漫符.
計	28	

#### 4.2.1 実験 1-1: 再現性 個人試行間一致

個人内でも画像に対する印象は変わりやすく、試行間で付与するタグに異なりがある可能性が高い。個人内で印象が変わらない画像にはどのような特徴が見られるか調べるため、各データセットに対し、タグ付けを日を分けて 2 回ずつ実施した。各データセット  $q$  に関する 1 回目と 2 回目に割り当てられたクラスの一致率を以下のように定義する。

$$\frac{\sum_{i=1}^N M(C(s_i^a, 1), C(s_i^a, 2))}{N} \quad (1)$$

表 1 に試行間の一致率を示す。漫画特徴の特徴的要素が多い方が一致率が高いことがわかる。

#### 4.2.2 実験 1-2: データセット間的一致

3 通りの組合せのデータセット間的一致率を調べた。 $a, b$  のデータセット間の  $t$  回目の一致率を次のように定義する。

$$\frac{\sum_{i=1}^N M(C(s_i^a, t), C(s_i^b, t))}{N} \quad (2)$$

表 2 にデータセット間的一致率を示す。1 回目も、2 回目においても、データセット間での一致率の大小関係は同じであり、漫画の特徴的要素の差の大小関係と対応した。

### 4.3 考察

表 3 に全クラスと頻度を示す。以下、4.2.1 でいずれのデータセットにおいても試行間で一致したデータのみに関して、データセット間的一致パターン全 5 種について考察する。表 4 に一致種別ごとの一致率を示す。なお、それぞれの一致パターンの画像には以下の特徴があるといえる。

全一致 画像特徴が印象に関連。印象が明確。

#### 部分一致

オリジナル画像と台詞なし画像 ( $a = 1, b = 2$ ) 吹き出し形状が印象に関係。

オリジナル画像と吹き出し統一画像 ( $a = 1, b = 3$ ) 吹き出し形状と台詞内容が一致せず、吹き出し形状がノイズとなることが想定される。

台詞なし画像と吹き出し統一画像 ( $a = 2, b = 3$ ) 台詞内容が印象に関連

#### 全不一致 印象が変わりやすい画像

表 5 に全一致のクラスと頻度、コマ内の画像特徴を示す。登場人物間の表情、顔周りの漫符、登場人物間の感情関係が重要な特徴となっていることがわかる。全一致したクラスが今回のアノテータの中で一貫したクラスと考えられる。今回はなるべく自由度を高くして検討するため実験間でクラスを固定しなかったことから、重複が少なくなったが、今後、実験人数を増やすか、あらかじめ複数人に固定したクラス候補集合からクラスを割り当てるように指示する。全一致の画像は台詞文字列や台詞形状などの台詞オブジェクトの影響に左右されず、画像特徴のみで理解できると言える。

## 5. 実験 2: 計算機によるクラス分類

畳み込みニューラルネットワーク (CNN:Convolutional Neural Network) は画像の形状特徴に強く反応する識別器を構築する

表 3: 全クラス頻度表

オリジナル ( $d=1$ )				台詞なし ( $d=2$ )				吹き出し統一 ( $d=3$ )			
1 回目 ( $t=1$ )		2 回目 ( $t=2$ )		1 回目 ( $t=1$ )		2 回目 ( $t=2$ )		1 回目 ( $t=1$ )		2 回目 ( $t=2$ )	
頻度	クラス	頻度	クラス	頻度	クラス	頻度	クラス	頻度	クラス	頻度	クラス
57	呆れ	47	呆れ	32	呆れ	26	困惑	59	呆れ	44	呆れ
23	悲しみ	26	悲しみ	22	驚き	25	呆れ	28	困惑	40	どんまい
20	驚き	21	驚き	21	悲しみ	23	驚き	20	ほのぼの	20	ドン引き
18	ほのぼの	17	怒り	20	困惑	21	悲しみ	18	驚き	18	ほのぼの
14	怒り	14	納得	18	怒り	20	怒り	16	怒り	14	怒り
10	どんまい	13	ほのぼの	17	ほのぼの	20	喜び	16	どんまい	13	困惑
8	喜び	10	ドン引き	14	喜び	14	ほのぼの	9	悲しみ	12	驚き
7	前向き	9	困惑	12	ドン引き	12	不安	8	浮かれ	5	否定
7	困惑	8	前向き	9	不安	11	ドン引き	5	焦り	5	諦め
5	恐怖	7	どんまい	9	どんまい	10	どんまい	4	恐怖	4	余計な一言
5	ドン引き	6	言い訳	7	恐怖	4	恐怖	3	余計な一言	4	反省
4	納得	4	恐怖	4	焦り	2	しみじみ	2	やけくそ	3	恐怖
4	心配	3	喜び	2	諦め					3	喜び
2	言い訳	1	不安	1	否定					2	自業自得
1	浮かれ	1	逃げ							1	浮かれ
1	反省	1	心配								
1	恥ずかしい										
1	勘違い										

表 6: 学習パラメータ

Batch size	20
The number of epoch	100
Dropout ratio	0.5
Activation function	ReLU function
Loss function	Softmax cross entropy
Optimizer	Adam. default parameter

表 8: プーリング層のパラメータ

	POOLING1	POOLING2
Size	2 × 2	3 × 3
Padding size	0	0
Stride	2	3

表 7: CNN 層のパラメータ

	CNN1	CNN2
Filter size	5 × 5	5 × 5
Padding size	0	0
Stride	1	1

際に有用であることが知られている。そのため、CNN では画像特徴をよく識別でき、各データセットを与えて学習させ、テストデータの画像特徴がどの程度、画像特徴由来かを調べることができると考えられる。その際の識別失敗データは画像特徴以外の特徴を用いて理解すべきシーンであると仮定できるため、失敗データを分けて考察することで、漫画理解に必要な他の特徴を検討することができる。筆者の前実験により、白黒画像の 4 コマ漫画の識別にも CNN が有用であることを確かめている。

本実験では 4. で人が割り当てたクラスを正解とする。学習に Python フレームワークの Chainer を用いた。表 6 にニューラルネットワークのパラメータを、表 7, 8 に畳み込み層とプーリング層のパラメータをそれぞれ示す。ネットワークの構成は以下の通りである。実験結果は発表時に示す。

#### 層の構成:

Input - CNN1 - ReLU - MAX POOLING1 - CNN2 - ReLU - MAX POOLING2 - LINEAR1 - ReLU - Dropout - LINEAR2 - Output

## 6. まとめ

本研究では、漫画の理解過程を工学的に実現するため、4 コマ漫画の 4 コマ目の特徴的要素とストーリーの印象の関係性を調べた。その結果、登場人物間の感情を登場人物の表情と吹き出し形状から理解することが有用である可能性が高いことを示した。今後の課題は以下の通りである。

- 複数作品、複数被験者による実験により、コマクラスをベクトルとして表し、要素を頻度とし、共感されやすいクラスを把握した上で個性を考慮する。
- コマ間のコマクラスの変化と 4 コマ全体のストーリー印象の関係を検討する。

## 参考文献

- [1] 松下光範. コミック工学の可能性. 第 2 回 ARG WEB インテリジェンスとインタラクティブ研究会, pp. 63–68, 2013.
- [2] A. Krizhevsky, I. Sutskever & G. E. Hinton, *Imagenet classification with deep convolutional neural networks*, In *Advances in neural information processing systems*, pp. 1097–1105, (2012)
- [3] 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.