4F1-5in2

漫画中の表現獲得方法に基づくストーリー理解過程の解析

An Analysis for Interpretation Process of Stories based on Comics Features

上野 未貴 *1　末長 寿規 *1　井佐原 均 *1
Miki Ueno　Toshinori Suenaga　Hitoshi Isahara

*1豊橋技術科学大学
Toyohashi University of Technology

In the artificial intelligence field, four scene comics is the suitable target for interpretation process of stories. This paper describes two kinds of our proposed classification; to classify emotion of comic scenes and to classify the role of order of four scene comics. Based on these results, we propose several equations to estimate story patterns and caption generation of four scene comics as an example task of interpretation process of stories.

1. はじめに

計算機によるストーリー解析 [1] や自動生成 [2] の研究は小説，すなわち文章を主な対象として，自然言語処理分野の手法を用いて取り組まれてきた．一方で，絵と文章が相補的な関係にある媒体として，絵本の自動生成 [3] や工学分野における漫画の研究 [4] に漫画のコマ画像の領域検出 [5]，漫画のレビュー分析などがある．

筆者らは漫画のストーリーの特徴設計に着目し，以下の 2 つを主な理由として，計算機手法との親和性の高さから，4 コマ漫画を対象として研究を進めている．

- 話の長さおよびコマの画像の大きさが統一されている
- 話の起承転結が他の媒体に比較して明瞭な傾向にある

研究の最終目的は，漫画の全体的なストーリーの自動生成および未知の漫画の画像を入力として説明文を自動生成するなどストーリーの意味解析の過程を工学的に実現することである．これまで，2 コマ漫画の自動生成システムの提案 [6] やストーリーの理解を一種の識別問題と捉えてタスクを設計し，人と深層学習による 4 コマ漫画の感情識別 [7] や順序識別 [8] に寄与する画像および言語特徴の検討を進めた．本稿では，識別時に得られた特徴を整理し，各コマの感情と順序のクラスの組合せによるストーリーのパターンを定義し，ストーリーパターンの推定およびキャプション生成を計算機によるストーリー理解過程の解析の具体的目標として位置づける．

2. 感情識別

本章では，筆者らが提案した漫画の画像の感情識別 [7] に関して詳述し，人手の識別結果と計算機実験の結果を比較して，感情識別に必要な特徴を考察する．

2.1 データセット

「こんぺいと!1[9]」の漫画の 1 巻の 188 話の 4 コマ目のみを使用して感情識別に関するデータセットを作成した．画像サイズは 290 × 210 ピクセルである．漫画は画像と文字特徴の複合的な媒体であるため，画像と文字特徴のいずれが重要か考慮するため，コマの一部を恣意的に隠した 3 種のデータセットを構築した．図 1 に 3 種のデータセットを示す．

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

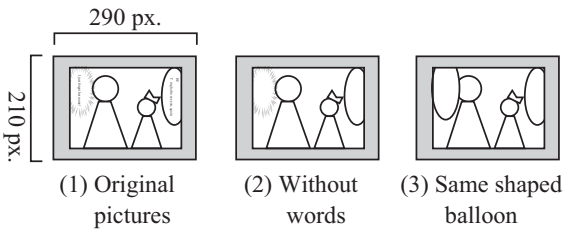

図 1: Each example of three types dataset

2.2 感情クラスの定義

同じ被験者でも時間によって回答が変わる可能性を考慮して，被験者 (20 代男性 1 名) に 3 種のデータセットについて各 2 回，6 日間を要して以下の手順でタグを付与させた．

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．

2.3 感情クラスの定義の結果

6 日間の実験で得られた結果は既発表 [7] の表 3 に示している．結果から以下が明らかになった．

- 被験者は画像特徴，特に漫符によって，コマの感情を識別する傾向が強いことが示唆された．
- 本実験では，1 つのコマ全体に対して 1 つの感情クラスを付与させたが，同じコマ内に複数のキャラクタが出現するときには，各キャラクタの感情を付与して，複数キャラクタの感情の組合せを考慮することが望ましい．
- 人がコマから受ける感情として，どのような文字列を想起するのか調べるために自由記述としたために，重複するタグが少なかった．人の多様性は考慮する余地を残したいものの，計算機実験でクラスを定義してクラスに共通する特徴を学習するためには，各クラスに属す一定量のデータが必要となる．従って人が想起する自由な記述を，指定する感情クラスに対応付ける必要がある．また，感情クラスの候補集合をあらかじめ被験者に与えて選ばせる，複数の感情のベクトル表現で表すなどが望ましい．

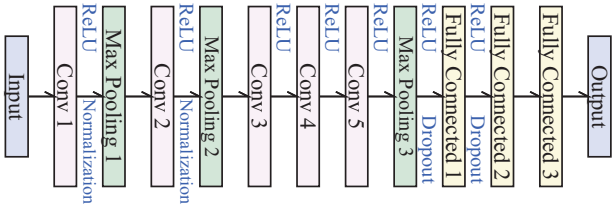

図 2: The architecture based on AlexNet[11]

表 1: Computer spec

OS	Ubuntu 14.04 LTS
GPU	GeForce GTX TITAN X(12 GB) × 2
CPU	Intel(R) Core(TM) i7-5930K CPU @ 3.50 GHz
Memory	32 GB
Chainer	1.10.0
Java	1.8.0_91
Python	2.7.6

2.4 識別実験に使用する画像とクラスの定義

1 名の被験者の感情クラスの特徴を深層学習で識別可能か調べるため，2.3 の結果の中で，各データセットで 2 日間試行で付与された感情クラスが共通の画像のみを使用し，付与された感情クラスを暫定的に正解ラベルとして計算機実験をした．被験者が回答した自由記述のクラスを，第 2 筆者が基本 6 感情 [10] に対応付けた．

6 クラスに対応付けた結果，枚数が比較的多かった，悲哀 30 枚，嫌悪 44 枚を使用することとした．

背景画像がコマ中に占める割合が大きいため，背景の模様が特徴的なコマは識別が容易であると考え，背景以外の特徴について詳細な考察をするため，各クラスから背景模様が広く含まれる画像を取り除き，以下のデータを用いて感情に関する 2 クラス識別をした．

1. 悲哀 18 枚
2. 嫌悪 22 枚

2.5 被験者による識別実験

以下のように，交差検定の手順を模して，人の漫画画像の感情の識別率を調べた．

1. 2.4 で使用することを決めた「悲哀」「嫌悪」クラスの画像を，訓練データとテストデータにランダムに 各クラス 3:1 の比率を保つよう分割した．訓練データは 28 枚，テストデータは 12 枚である．
2. 被験者 (20 代，男性 2 名）に訓練データを 10 分間見せて，特徴をメモさせた．
3. 訓練時にメモした特徴を参考にしながら，テストデータに，いずれかの感情クラスを回答させた．

2 名の識別率はいずれも 0.83 で，正解画像はすべて共通だった．2 名がメモした各クラスに共通する特徴は,「悲哀」は「泣いている」,「嫌悪」は「汗マーク」「半目」である．特に，2 つのクラスを識別するため，ため息，青線などの漫符や，目の開き具合などの細かい表情に着目していた．

2.6 計算機による識別実験

人の識別実験と同じデータを用いて，畳み込みニューラルネットワーク (CNN:Convolutional Neural Network) を構築して，感情クラスの識別率を調べた．図 2 にネットワーク構造を，表 1，2，3，4 に計算機のスペック，ニューラルネットワーク，プーリング層，畳み込み層のパラメータをそれぞれ示す．4 分割交差検定の平均識別率は 0.70 となった．

表 2: Parameters of neural network

The number of output units	2
Batch size	20
Node size of fully connected layer 1	256
Node size of fully connected layer 2	256
The number of epochs	600
Dropout ratio	0.5
Activation function	ReLU function
Loss function	Softmax cross entropy
Optimizer	Adam:alpha=1e-6

表 3: Conv layer parameters

	Conv 1	Conv 2	Conv 3	Conv 4	Conv 5
Filter size	5 × 5	5 × 5	5 × 5	5 × 5	5 × 5
Padding size	0	2	1	1	1
Stride	2	1	1	1	1

表 4: Pooling layer parameters

	Pooling 1	Pooling 2	Pooling 3
Size	3 × 3	3 × 3	3 × 3
Padding size	0	0	0
Stride	2	2	2

3. 順序識別

本章では，筆者らが提案した漫画の画像の順序識別 [8] に関して詳述し，人手の識別結果と計算機実験の結果を比較して，順序識別に必要な特徴を考察する．

3.1 データセット

「こんぺいと!1[9]」の漫画の 1 巻の 188 話を使用して，9 種類の 2 クラス識別に関するデータセットを作成した．図 3 に，データセットの作成手順を示す．画像サイズは，Set 1 は， 227 × 157 ピクセル，Set 2 は 227 × 314 ピクセルである．

Set 1-(n, m) において，n 番目のコマはクラス 1 に割り当てられ，m 番目のコマはクラス 2 に割り当てられる．

Set 2-(i-j, k-l) において，i 番目と j 番目のコマの組み合わせ画像はクラス 1 に割り当てられ，k 番目のコマと l 番目のコマはクラス 2 に割り当てられる．また，Set 2 の各画像は，2 コマの画像を垂直方向に連結して作成したため，高さは 1 コマの 2 倍であり，幅は 1 コマに等しい．

各クラスの画像枚数は 188 枚である．

3.2 被験者実験の手順

被験者は 20 代の男性 4 名である．

1. 上記の 9 種類の各 2 クラスのデータセットから各 10 枚の画像をランダムに抽出する．
2. 被験者に，画像のコマ番号を各 2 クラスのいずれかから選択させ，理由を回答させる．

3.3 被験者実験の結果

表 5 に，順序識別の人手実験の結果を示す．結果から，1，4 コマの 2 クラス識別は容易であり，コマを特徴づける要素に大きな差があることが示された．話の導入部らしさ，話の結末らしさ，といった特徴はコマの画像特徴や台詞特徴から抽出できる可能性が明らかになった．

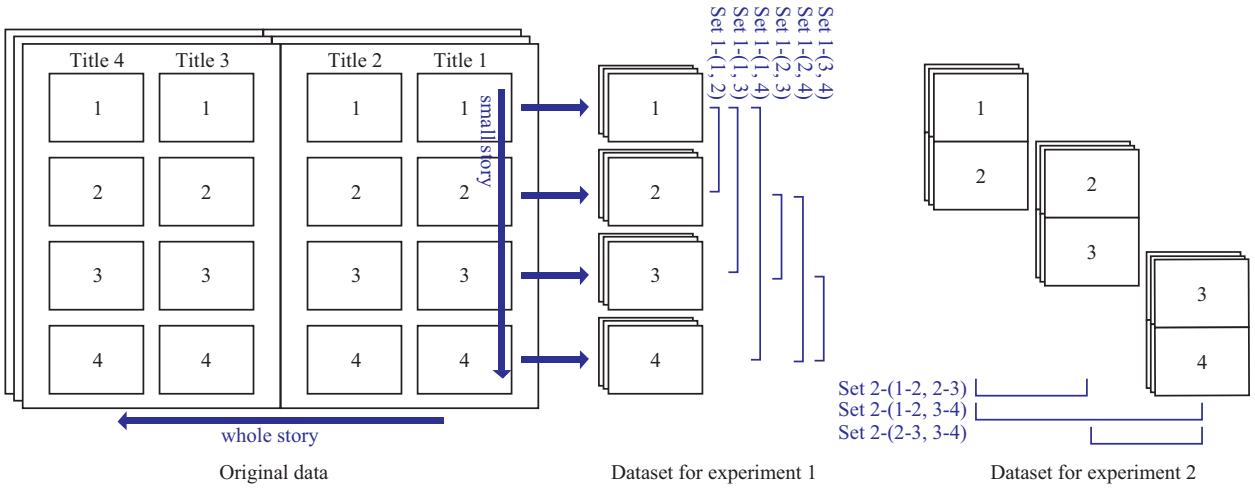

図 3: The Process of Preparing Experimental Dataset

表 5: Accuracy of Subjects

	識別率				
Set	User1	User2	User3	User4	平均
Set 1-(1, 2)	0.70	0.80	0.80	0.40	0.675
Set 1-(1, 3)	0.70	0.80	0.90	0.70	0.775
Set 1-(1, 4)	1.00	1.00	0.80	0.90	0.925
Set 1-(2, 3)	0.80	0.50	0.80	0.70	0.700
Set 1-(2, 4)	1.00	0.70	0.90	0.80	0.850
Set 1-(3, 4)	0.90	0.70	0.70	0.70	0.750
Set 2-(1-2, 2-3)	0.80	0.80	0.60	0.80	0.750
Set 2-(1-2, 3-4)	0.90	0.90	0.80	0.80	0.850
Set 2-(2-3, 3-4)	0.70	0.60	0.90	0.80	0.750

表 6 に，順序識別時に被験者が付与した自由記述の回答理由の意味を考慮し，理由種別を分類してまとめた結果を示す．

3.4 計算機実験の結果

人手実験と同様のデータを用いて畳み込みニューラルネットワークにより順序識別をした [8]．結果，Set 1 に関しては，Set1-(1,4) の識別率が最も高く，500 エポックで，4 分割交差検定を 5 回実施した平均識別率は 0.62 となった．

4. 応用: 4 コマ漫画のストーリーを表すキャプション生成

深層学習の発展により，写真画像のキャプション生成 [12] の発展が目覚ましい．静画のキャプション生成や，限定的なオブジェクトしかない監視対象の室内の行動や限定的な行動しかとらないスポーツなどについては，従来から動画，すなわち連続したシーンのキャプション生成の研究が存在する．一方で，現実世界の常識に当てはまらないあらゆる事象が起こり得る漫画のキャプション生成は，ストーリーパターンの多様性，写真画像に比して難しいとされる白黒のイラストの識別，という観点から，挑戦的な課題 [13] である．

本稿ではストーリー理解を一種の識別問題と捉えることで単純化したが，本章では，その識別問題の応用について述べる．すなわち，計算機の漫画理解の一つの実現例として，4 コマ漫画の画像を入力として，上記で述べた 2 種の識別を例として，各話のストーリーのキャプションを出力する方法を検討する．

∗1 Set 2 でのみ見られた理由

表 6: Comments of Order Classification

種別	説明	被験者の自由記述の例
直観	ただの直観	直観
台詞依存	会話文やナレーションなどコマ中の台詞を理由とする	～という発言から
「起」	起承転結の「起」	話の導入っぽい
「転」	起承転結の「転」	話の転換点っぽい，
「結」	起承転結の「結」すなわち対象コマがオチか否か	「オチっぽい」
予測	上下の他コマ内容を予測して，対象コマを位置づけ	このあと何か起こりそう，このあと何か言いそう，前のコマがなくてもコマとして成立する，コマ内容が唐突
消去法	他の理由クラスと対比	転換点ではない
比較 ∗1	上下コマを比較する	1 コマ目が「起」にならなそう
その他	上記以外	

4.1 順序識別とストーリーパターンの定義

順序識別からコマ間の意味的な距離を算出する．同じ話の 2 クラス 6 種の Set 1 のデータについて，順序識別の識別率をベクトルで表し，4 コマ漫画で典型的なストーリーパターンとの相関を調べる．Set 1-(n, m) の識別率を $a_{n,m}$ とし，6 つの要素から成る実数値ベクトル $\boldsymbol{s}_{\mathbf{order}}$ を以下のように定義する．

$$s_{\text{order}} = (a_{1,2}, a_{1,3}, a_{1,4}, a_{2,3}, a_{2,4}, a_{3,4}) \tag{1}$$

上記のベクトルの要素の値を用いて，日本の 4 コマ漫画の典型的なストーリーパターンの特徴と数式表現を例示する．

起承転結 導入と結末が明確．

$$(1, 4) = \arg\max_{n,m}(a_{n,m}) \tag{2}$$

出オチ 最初のコマが最も特徴的．

$$\sum_{n=2}^{4} a_{1,n} > a^{\text{threshold}} \tag{3}$$

繰り返し 1-2 コマと 3-4 コマで類似事象が起きる．3-4 コマの事象の方が強調される．

$$|a_{1,2} - a_{3,4}| < a^{\text{threshold}} \tag{4}$$

4.2 ストーリーキャプション生成

2. で任意のコマの感情クラスの識別について述べた．以下のように，n コマ目の感情クラス e_n の系列ベクトルを定義して，ストーリーパターンとの相関を調べる．

$$s_{\text{emotion}} = (e_1, e_2, e_3, e_4) \tag{5}$$

上記の感情クラスベクトルと順序ベクトルによって，4 コマ漫画の各話を特徴づける系列ベクトルとストーリーの意味的相関を調べ，単なる 4 コマ漫画画像を入力として，ストーリーのキャプション生成をする．ストーリーのキャプション生成のために，上記のベクトルから，特定のストーリーを表すキャプションのテンプレートを選択する．以下に具体的なキャプションのテンプレートの例を挙げる．

出オチ 突然，[Object1] が [Action1] する話．
繰り返し [Object1] が [Action1] した後，さらに，[Object2] が [Action2] する話．

空欄補充が必要な部分を，[Object1] というようにオブジェクト名で示している．生成したいキャプションの具体的な例を以下に示す．

出オチ 突然，キャラクタ 1 が叫んでも誰にも反応してもらえずに寂しい思いをする話．
繰り返し キャラクタ 1 が家で忘れ物をした後，さらに，キャラクタ 1 が電車の中で忘れ物をする話．

漫画のオブジェクト認識，オブジェクトの変化や感情の変化に基づいて，オブジェクトを表す単語や感情を表す単語を決定し，テンプレートの空欄を補充して，ストーリーを表すキャプションを生成する．

漫画のストーリーを過不足なく文で表すことは人にとっても困難だが，イラスト認識，識別の精度向上と，ストーリーの上記ベクトルの定義と出力を望むキャプションの形式を対象作品と被験者実験の人数を増やして検討することで計算機上で実現可能と考え，発表時にデータを示して再度検討する．

5. 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 コマ漫画のストーリーの理解過程を感情識別と順序識別に帰着して，識別に寄与する特徴を考察し，ストーリーのパターンを数式表現で定義する方法を提案し，ストーリー理解を具体的なタスクとしてキャプション生成と位置付けることを示唆した．今後の課題は以下の通りである．

- 複数被験者への追実験から個性の影響を考慮した感情クラスの再定義
- 複数作品を対象とした感情識別と順序識別の追実験からストーリーパターンの変数と閾値を決定
- 4 コマ漫画のストーリーのキャプション生成のためのデータセットの構築

謝辞

本研究は一部，大幸財団の支援による．

参考文献

[1] Vladimir Iakovlevich Propp. *Morphology of the folktale*. University of Texas Press, 1968.

[2] Terry Winograd. Procedures as a representation for data in a computer program for understanding natural language. *Cognitive Psychology*, Vol. 3, No. 1, pp. 1–191, 1972.

[3] 福田清人, 藤野紗耶, 森直樹, 松本啓之亮. ストーリーモデルによる絵を用いたシナリオ生成. 人工知能学会, No. 2J5-OS-08b-3, 2016.

[4] 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ション研究会, pp. 63–68, 2013.

[5] Yusuke Matsui, Kota Ito, Yuji Aramaki, Toshihiko Yamasaki, and Kiyoharu Aizawa. Sketch-based manga retrieval using manga109 dataset. *CoRR*, Vol. abs/1510.04389, , 2015.

[6] 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.

[7] 上野未貴, 森直樹, 松本啓之亮. 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. 人工知能学会, No. 2J5-OS-08b-4in2, 2016.

[8] Miki Ueno, Naoki Mori, Toshinori Suenaga, and Hitoshi Isahara. Estimation of structure of four-scene comics by convolutional neural networks. In *Proceedings of the 1st International Workshop on coMics ANalysis, Processing and Understanding, MANPU@ICPR 2016*, pp. 9:1–9:6, 2016.

[9] ふじのはるか. こんぺいと!1. 芳文社, 2007.

[10] Paul Ekman and Wallace V Friesen. Constants across cultures in the face and emotion. *Journal of personality and social psychology*, Vol. 17, No. 2, p. 124, 1971.

[11] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[12] Andrej Karpathy and Li Fei-Fei. Deep visual-semantic alignments for generating image descriptions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3128–3137, 2015.

[13] 紀子新井, 拓也松崎. ロボットは東大に入れるか? : 国立情報学研究所「人工頭脳」プロジェクト (¡特集¿ロボットは東大に入れるか?). 人工知能学会誌, Vol. 27, No. 5, pp. 463–469, sep 2012.