

コマの持つ属性を用いたマンガのシーン自動抽出

Automatic extraction of scene of Manga using the attribute of frames

久行 智恵^{*1} 三原 鉄也^{*2} 永森 光晴^{*3} 杉本 重雄^{*3}
Chie Hisayuku Tetsuya Mihara Mitsuharu Nagamori Shigeo Sugimoto

^{*1} 筑波大学情報学群情報メディア創成学類
College of Media Arts, Science and Technology University of Tsukuba

^{*2} 筑波大学図書館情報メディア研究科
Graduate School of Library, Information and Media Studies University of Tsukuba

^{*3} 筑波大学図書館情報メディア系
Faculty of Library, Information and Media Studies University of Tsukuba

Readers of a Manga first see frames in a single page, recognize the order of frames in the page, read the pages of the Manga and then understand its story. They identify a scene of a story as a single entity delimited in a temporal line of the story. Extracting scenes is beneficial to create an abstract and an index of a Manga. This paper shows an experimental automated extraction of scenes from a Manga based on appearance and layout of frames.

1. はじめに

デジタル環境でのマンガの流通や利用が広く普及した現在、内容や構造に即したより高度なマンガコンテンツの利用が期待される。ストーリーはマンガの主要な内容の情報のひとつであり読者はマンガコンテンツからそれを読み取り解釈しているが、マンガを構成する画像データには明示的に記述されていない。また大量のマンガ画像について人手でその構成情報を作成することは困難である。従ってマンガのストーリーに関する情報のデジタル上での利用は進展していない。

こうした状況に対し、本研究ではストーリーの最も基礎的な構成単位であるシーンに着目し、コマの持つ属性について記述したメタデータに対して機械学習し、シーンの転換箇所となるコマを自動的に判定することでシーンの抽出を試みた。まず複数のマンガ作品について人手でシーン転換箇所となるコマの判定とコマの持つ属性についてのメタデータの作成を行う。このデータを対象に、コマに含まれるマンガ表現である、キャラクターやテキスト情報の有無、コマそのものの位置や形状を特徴量としてSVM(Support Vector Machine)による機械学習を行い、シーン転換箇所の判定実験を行う。さらに、その実験結果からシーンの自動抽出の有効性や効率的な学習の条件について考察する。

2. マンガにおけるシーンとその抽出

シーンとは本来映像についての概念である。映像におけるシーンとは「物語映画の一つの単位(ユニット)」であり、「連続した時間・空間において起こり、時間・空間のいずれか、あるいは両方の変化とともに終わる」ものである[Blandford 04]。映像は画像を連続して提示することでストーリー・場面を表現するものである一方、マンガも平面上に複数のコマを配置することでストーリーを表現するものである。複数の画像の集合でストーリーを表現

する点においてマンガは映像と同様であり、従ってマンガにおけるシーンも映像におけるシーンと同様に捉えることができる。よって本研究では、マンガの1コマは映画の1カットと同等であり、同一の時間、同一の場所のコマの連続をひとつのシーンとみなす。シーンを特定することで、紙媒体のようにページをパラパラと捲ってストーリーの大まかな内容を確認することが難しいデジタルマンガにおいても、重要シーンの抽出によるダイジェスト作成やシーン先頭のインデクシングが可能となり、読者は既読作品を効率的に探索できる。しかしシーンはマンガ画像に明示的には記載されていない。マンガ画像からその内容に関する情報を識別する研究は既にあるが[野中 12][石井 13]、これらはマンガ画像の画像的特徴を用いたオブジェクト認識による方法であり、情報の識別はオブジェクト認識の精度に大きく依存する。加えてマンガ画像のオブジェクト認識の技術は十分に確立されていない。また、マンガコンテンツは単一の作品についても多くの画像から構成されるため、マンガ画像からシーンを抽出するためには大量の画像を扱う必要がある。これらの理由により、マンガ画像からシーンを直接抽出することは困難であると考えられる。

そこで筆者らは汎用的なメタデータとしてMMF(マンガメタデータフレームワーク)に基づくメタデータを利用し[三原 15]、マンガの意味内容が反映される構成要素をコマの特徴量として用いることでシーンが転換するコマの抽出を試みることを考えた。MMFとはマンガの発見やアクセス、再利用を目的とした、書誌記述、構造記述、知的内容に関する記述の観点に基づくメタデータ記述モデルである。筆者はそのMMFに基づいてシーンの最初となるコマの構造記述と書式術に関するメタデータ要素の分析を行った[久行 15]。シーン最初のコマでは、読者に場面の説明のための空間・時間的な奥行き表現がしやすい縦横に長いコマを用いて背景描写が行われることが多い。またページの捲りに合わせてシーンを変える手法も多々見られる。これらを受けて、シーン最初のコマにおけるキャラクターやテキスト情報の有無、コマの形状、ページ内におけるコマの位置などのMMFで記述可能なコマの属性のパターンについて調査分析し

連絡先: 久行智恵, 筑波大学情報学群情報メディア創成学類, 〒305-8550 茨城県つくば市春日 1 丁目 2, s1311445@u.tsukuba.ac.jp

た。加えて作品のページ数や著者情報も要素とし、これらの要素を条件として組み合わせることでシーンの抽出が可能であることを明らかにした。

しかしながら文献[久行 15]ではシーン転換コマに関する各条件を手作業で検証しているが、より多くのマンガについてシーンの利用を行うためには自動的な抽出が不可欠である。そこで本研究では、マンガのコマ及びその属性を特徴量とした機械学習によるシーン転換コマの判定を実験的に行う。その結果からシーン抽出に有効な特徴量の検討を行う。

3. 関連研究

映像を対象にしたシーンの自動抽出に関して、山内や関口らの研究がある[山内 14] [関口 06]。山内は動画に付与された視聴者のコメントを利用した特徴的なシーンの抽出手法を提案している。マンガにおいてもニコニコ静画^{※1} やぽこぽこ^{※2} のようにコメントが付与される作品もあるが、特定のサービスにおいてのみ実現されている機能で、一般的であるとは言えない。関口は野球中継映像の盛り上がりシーンの検出について、得点表示などのようなシーンごとに確実に変わるテロップや観客の歓声を利用する手法を提案している。マンガにはシーンの転換を確実に判断可能な要因は知られておらず、音声はマンガ画像には存在しない要素であるため、これらの手法はマンガに対して適用することはできない。

マンガの内容構成の抽出に関する研究として、藤岡はマンガを自動的に要約するための要約手法を提案している[藤岡 15]。藤岡はマンガ内容の重要部分を抽出し 30 ページ程度にまとめたものを作成した。5 ジャンル 5 作品で読みやすさとわかりやすさの観点から加点式で被験者に評価させる検証を行っており、重要部分の抽出はキャラクター毎の登場回数とセリフ数、コマの大きさの調査結果に合わせて作品によって異なる手法を用いた。結論として、作品内でコマの大きさが大きく変わる作品では 1 ページ使用したコマを中心に盛り上がった場面を含むシーンを抜き出した時に、キャラクター同士の会話などのようなセリフ表現が多い作品では話数毎に均等にページを抜き出した時にそれぞれ被験者から多くの加点を得ている。本研究においてもコマの属性に着目することでマンガ内容の抽出を検討する。コマの中に表現されるキャラクター、テキストの有無やコマの大きさ、形状、位置のようなコマの要素を特徴量として機械学習による自動的な抽出を試み、著者という書誌情報の観点から考察を行う。

4. 正解シーン転換箇所の設定

2 章で述べたマンガのシーンの定義に則り、本研究では作品中の時間や場所におけるコマの連続性が途切れた箇所をシーンの転換箇所と定義する。しかしシーン転換箇所について明示的な表現がなされることは稀であり、シーン転換箇所の識別は読者の感覚に基づく認知によって行われている。加えて、シーン転換箇所に関するマンガ表現技法についての議論も見られない。さらに、シーン転換箇所には読者の認知に基づくものの他、著者の意図という観点もある。著者が意図したシーン転換箇所と読者が認知する箇所が異なることもあり得る。

本研究ではシーンの自動抽出に先立って正解となるシーン転換箇所を設定する必要がある。そこでまず、シーン転換箇所における読者の認知と著者の意図の差異について検証した。木野陽(以下、著者 α) 著「マンガジュニア名作シリーズ 銀河

表 1 シーン転換箇所と記述された箇所の数

作品名	著者	読者 1	読者 2	読者 3	設定 A	設定 B	適合率 (設定 A)	適合率 (設定 B)
作品 C	82	73	50	71	62	67	42 (0.51)	52 (0.63)
作品 E	85	82	47	37	48	53	43 (0.51)	50 (0.59)

表 2 検証作品

著者	作品	ページ数	コマ数	シーン転換数
著者 α	作品 A	39	191	9
	作品 B1	35	205	12
	作品 B2	50	240	15
	作品 C	262	1333	62
著者 β	作品 D1	46	198	23
	作品 D2	32	139	10
	作品 E	203	960	48

鉄道の夜 (ISBN: 978-4052039201) (263 ページ、以下作品 C) と布袋あずき(以下、著者 β) 著「マンガジュニア名作シリーズ 小公女 (ISBN: 978-4052035012) (204 ページ、以下作品 E) の 2 作品について読者 3 名と著者それぞれにシーン転換箇所となるコマを指定させた。読者の指定するシーン転換箇所については被験者間での判断の揺れを考慮し、2 名以上の読者がシーン転換箇所と指定したコマを合議的なシーン転換箇所とした(設定 A)。さらに、読者の指定するシーン転換箇所について、同一のシーン転換について 1 コマ程度ずれて指定している場合が少なからずあった。そこでこの揺れを緩和するために、前後 1 コマ以内に 2 名以上がシーン転換箇所と記述されたコマを抽出し、連続したコマが抽出された場合はその連続したコマをひとつかたまりのシーン転換箇所とし(設定 B)、設定 A・B について著者が設定したシーン転換箇所と比較した。

表 1 は著者および各読者、設定 A・B それぞれのシーン指定箇所数と設定 A・B について著者が設定したシーン転換箇所に対する適合率を示したものである。設定 A の適合率は作品 C と作品 E とともに 0.51 であった。設定 B と著者が設定したシーン転換箇所との一致数は、設定 A に比べ作品 C は 10 コマ、作品 E は 8 コマ増え、設定 B の著者に対する適合率は作品 C が 0.63、作品 E は 0.59 となった。

読者および作者それぞれのシーン転換箇所設定数に大きな差は見られなかったが、約 5 割の箇所が一致していないことから、シーン転換箇所の自動抽出において、転換箇所の設定手法についても十分に議論される必要があると考えられる。

5. SVM を用いたシーンの自動抽出

本研究では、シーンを抽出するために教師あり機械学習 SVM を利用する。SVM は 2 クラス分類を行い、特徴量の次元が大きくなってもある程度識別精度が良いという特徴がある機械学習手法である。よって特徴量が複数存在するコマをシーン転換箇所か否かに分類するのに適した手法であると考えた。

筆者が行ったシーン最初のコマに対する分析を踏まえ[久行 15]、コマに関する要素を次のように変数化して特徴量とした。背景描写のみでキャラクターやセリフなど存在しないコマを抽出するために 1) コマの中にキャラクターが存在すれば 0、無ければ 1、2) コマの中にテキストが存在すれば 0、無ければ 1 とする。背景描写がなされることが多い縦横に長いコマを抽出するために 3) コマフレームの長辺が単辺の 4 倍以上あれば 1、無ければ 0 とする。またコマの形状において重要な要素である面積は 4) 1 ページを面積 1 とした場合のコマの相対的な面積として変数化した。ページの捲りに合わせてシーン転換が起こる場

※1 <http://seiga.nicovideo.jp/>

※2 <http://www.poco2.jp/>

合、その転換コマの位置はページの最初のコマであることが多いため 5) ページ内における相対的なコマの位置(コマ番号÷そのページにあるコマの数)とした。

6. シーンの自動抽出実験

6.1 実験条件

本実験では著者 α が自費出版した作品 A、同一タイトルの上下巻作品 B1・B2、作品 C の 3 タイトル 4 作品と、著者 β が自費出版した同一タイトルの上下巻作品 D1・D2、作品 E の 2 タイトル 3 作品を用いてシーンの自動抽出実験を行った。これら作品のすべてのコマについて、マンガメタデータ作成ツール ComicMetaEditor を用いて、作品名・ページ番号・ページ内におけるコマ番号・コマの頂点の位置情報・コマ内に描写された全てのキャラクター、テキスト(セリフ、ナレーション)情報といったメタデータ記述を作成し、5 章で述べた変数に変換し、それぞれのコマの特徴量とした。さらに正解シーン転換箇所の設定については、データ作成の容易さから、4 章で述べた設定 A によるシーン転換コマの設定を行った。表 2 はこれら作品のページ数、コマ数およびシーン転換箇所数を示す。

本実験の SVM による機械学習処理には Chang らの LIBSVM[Chang]を用いた。カーネルタイプは RBF カーネルを利用し、コストパラメータと C と RBF カーネルのパラメータ γ はグリッド探索で最適な組み合わせを算出することでチューニングした。

本実験では、各作品について教師データをサンプリングし、残りのデータをテストデータとしてシーン転換箇所の正否の2値分類を行った。教師データのサンプリング手法として、手法 1) 全てのコマを学習させる、手法 2) 全てのシーン転換コマと、それと同数の非シーン転換コマをランダムサンプリングして学習させる、手法 3) ある一定数(手法2と同数)をランダムにサンプリングして学習させる、の 3 つの手法について学習モデルを作成し、分類した時の適合率、再現率、F 値を求めて比較した。その結果、手法 1 と手法 3 については、手法 2 よりも F 値が高いモデルはいくつか作成できたものの、2値分類できない(全てシーン転換ではないと分類する)モデル作成されることがあった。これを踏まえ、本実験では全ての検証作品においてある程度の分類ができる手法 2 を教師データのサンプリング手法とした。

6.2 特徴量の検証

表 3 は各作品から作成した学習モデルの特徴量の重みと平均値、標準偏差を示したものである。5 章より特徴量はコマの各要素を変数化したものを用いており、1) character はキャラクターの存在に関する特徴量の重みを、2) text はテキスト情報の存在に関する特徴量の重みを、3) aspect はコマの形状に関する特徴量の重みを、4) area はコマの面積に関する特徴量の重みを、5) page_coma はページ内におけるコマの相対的な位置に関する特徴量の重みを示している。それぞれの特徴量の平均の最小値は page_coma の -89.83、最大値は area の 31.20 となった。標準偏差の最小値は text の 11.83、最大値は page_coma の 134.88 となった。

6.3 書誌的関連性に基づく学習モデルの検証

(1) 同一作品における学習モデルの検証

一つの作品中におけるシーン転換のコマの特徴は共通することが予想される。各検証作品を用いて作品全体と作品の前半のそれぞれから教師データを作成してモデルを作った時の分類精度を求めた。

表 3 各作品の学習モデルの特徴量の重み

モデル	1) character	2) text	3) aspect	4) area	5) page_coma
A	-63.80	16.78	156.84	-14.40	-392.12
B1	0.25	0.31	0.13	0.06	-0.05
B2	1.45	5.20	-2.69	8.68	-6.92
C	0.88	1.38	0.13	0.01	-0.92
D1	47.43	-10.15	-16.57	14.33	-70.17
D2	0.56	0.44	0.13	-0.10	-0.24
E	-22.48	28.76	26.17	209.83	-158.44
平均	-5.10	6.10	23.45	31.20	-89.83
標準偏差	30.82	11.83	55.70	73.39	134.88

表 4 同一作品内の学習モデルの精度 1

モデル	評価データ	Scene_true	Scene_false	適合率	再現率	F 値
b1	data_b1	12	12	0.10	0.40	0.16
b	data_b	14	14	0.08	0.38	0.13
d1	data_d1	23	23	0.09	0.90	0.16
d	data_d	17	17	0.24	0.56	0.33

表 5 同一作品内の学習モデルの精度 2

モデル	評価データ	Scene_true	Scene_false	適合率	再現率	F 値
a1	data_a1	5	5	0.03	0.75	0.06
a	data_a	5	5	0.06	1.00	0.11
c1	data_c1	36	36	0.03	0.77	0.05
c	data_c	32	32	0.03	0.60	0.06
e1	data_e1	25	25	0.10	0.95	0.18
e	data_e	25	25	0.08	0.83	0.14

作品 B1・B2 と作品 D1・D2 は1つのストーリーを書誌的に上下巻で分けたものであるが、これら上下巻の作品データを合わせたものを作品 B、D とする。作品 B1、D1 それぞれから 6.1 節で述べた手法 2 を用いて教師データを作成し、学習させなかったデータをそれぞれ作品 B2、D2 に加えて評価データを作成した。この際に作られたモデルをそれぞれ b1、d1 とし、評価データをそれぞれ data_b1、data_d1 とする。また作品 B、D それぞれから、各作品のシーン転換数の半分とそれと同数の非シーン転換コマを教師データとし、学習させなかったデータを評価データとした。この際に作られたモデルを b、d とし、評価データを data_b、data_d とし、それぞれ分類を行った。表 4 は各モデルにおける分類の適合率、再現率および F 値を示したものである。Scene_true と Scene_false は学習させたコマの数である。作品 B についてのモデルである b1 と b の F 値に大きな差は見られないが、作品 D のモデルである d1 と d を比較すると、モデル d は d1 よりも 0.2 ほど F 値が高かった。

作品 A、C、E は上下巻のように書誌的に分割されていない単一の作品である。これらの作品についてはページ数で作品を 2 分割した。作品 B、D と同様の手法でモデルと評価データを作成し、それぞれのモデルを a1、a、c1、c、e1、e とし、評価データを data_a1、data_a、data_c1、data_c、data_e1、data_e とした。表 5 はこれらのモデルで評価データを分類した結果である。適合率がどのモデルにおいても低く、F 値も全体的に低い。また、同一作品から作られたモデル同士の F 値に大きな差は見られない。

(2) 同一著者における学習モデルの検証

シーン転換コマの特徴は同一著者の作品と共通することが予想される。そこで各作品について 6.1 節で述べた手法 2 を用いて作成した教師データを学習させたモデルを、それぞれ A、B、C、D、E とした。学習で用いなかった作品それぞれの全てのコマを分類することでモデルを評価する。評価データは data_a、data_b、data_c、data_d、data_e とし、分類を行った。表 6 はそれぞれの分類における適合率、再現率および F 値を示したものである。著者 α のモデルで同一著者の作品を分類した時と異なる

表 6 同一著者における学習モデルの精度

モデル	著者 α									著者 β						
	data_a			data_b			data_c			data_c			data_e			
	適合率	再現率	F 値													
著者 α	A	-	-	-	0.14	0.82	0.24	0.13	1.00	0.23	0.08	0.55	0.14	0.09	0.65	0.16
	B	0.35	0.90	0.50	-	-	-	0.08	0.44	0.14	0.24	0.75	0.36	0.12	0.58	0.20
	C	0.09	0.65	0.16	0.13	0.58	0.21	-	-	-	0.18	0.63	0.28	0.25	0.40	0.31
著者 β	D	0.21	0.53	0.30	0.12	0.60	0.20	0.26	0.85	0.40	-	-	-	0.19	0.52	0.28
	E	0.13	0.75	0.22	0.10	0.64	0.17	0.09	0.79	0.16	0.23	0.90	0.37	-	-	-

著者の作品を分類した時の F 値の平均はそれぞれ 0.25 と 0.24、著者 β のモデルではそれぞれ 0.32 と 0.24 であった。

7. 考察

6.2 節では本実験における有効な特徴量について検証した。表 3 で示した特徴量の平均値と標準偏差の値の比較から、学習モデルによって重みの絶対値が大きくなる傾向があることが分かる。この結果から、シーン転換箇所の抽出モデルにおける重要な特徴量は各学習モデルによって異なると考えられる。

6.3 節(1)では同一作品における学習モデルについて検証した。表 4 より、上下巻として書誌的に分割されている作品 B はモデル b1 とモデル b の結果に大きな違いは見られなかったが、同じく分割されている作品 D ではモデル d の F 値がモデル d1 よりも値が高かった。この結果は同一作品内におけるシーン転換の特徴は似ていることから、学習のサンプリング箇所が異なっても F 値は変わらないという予測とは異なっている。これらは今回作成したモデルそのものの F 値が 1~3 割程度であることから、教師データの少なさに起因する分類精度の低さによるものだと考えられる。書誌的に分割されていない作品 A~E においても、表 5 の F 値の値が低いことから、書誌的に分割されている作品についての検証と同様であると考えられる。

6.3 節(2)では同一著者における学習モデルについて検証した。同一著者の作品を分類した時と、異なる著者の作品を分類した時の F 値の平均を比較すると、著者 α のモデルは、同一著者の作品を分類した場合の F 値の平均と異なる著者の作品を分類した場合の F 値の平均に大きな差は見られなかった。著者 β のモデルは同一著者の作品を分類した場合の F 値の平均の方が異なる著者の場合よりも良かった。しかし著者 α の作品 C から作られたモデルによる分類で最も精度が良いのは著者 β の作品 E であるように、異なる著者の作品から作られたモデルの分類精度が良いという結果も得られた。今回の実験では、同一著者の作品におけるシーン転換コマの特徴の共通性を示すことはできなかった。今後さらに多くの作品について検証する必要がある。

今回の実験では全体を通じて適合率が低い結果となった。これはシーン転換箇所についてのデータが少なく、十分な機械学習が行えなかったものと考えられる。これらを改善するためには、実験に用いる作品データを増やすことに加え、4 章に述べた設定 B を用いたシーン転換箇所の指定手法やシーン転換箇所の抽出範囲について検討する必要がある。

8. おわりに

本稿では、SVM を用いてシーン転換箇所を自動判定するモデルの作成を行い、書誌的観点からモデルを検証した。より精度の高いシーン抽出を実現するためには、単一のコマの抽出ではなく更に範囲を広げたシーン転換部分の抽出と評価を検

討したい。また、より多くの作品についての適用を議論するために、異なる著者や作品の検証作品を追加することは急務である。さらにより適切な特徴量の設定や SVM 以外の学習アルゴリズムの適用についても検討の余地がある。

謝辞 本研究の遂行にあたって実験用マンガ作品をご提供頂いた木野陽氏、布袋あずき氏、およびメタデータ作成ツールをご提供くださった石井大祐氏に深謝する。

参考文献

- [Blandford 04] Blandford, S., Grant, B.K., Hiller, J.: フィルムスタディーズ辞典 映画映像用語のすべて, フィルムアート社, 2004.
- [野中 12] 野中俊一郎, 沢野哲也, 羽田典久: コミックスキャン画像からの自動コマ検出を可能とする画像処理技術「GT-Scan」の開発, 富士フィルム研究報告, 57, pp.46-49, 2012.
- [石井 13] 石井大祐, 渡辺裕: マンガからの自動人物検出と識別に関する一検討, 画像電子学会誌, 42(4), pp.457-465, 2013.
- [三原 15] 三原鉄也, 永森光晴, 杉本重雄: マンガメタデータフレームワークに基づくデジタルマンガのアクセスと制作の支援—デジタル環境におけるマンガのメタデータの有効性の考察—, 電子情報通信学会論文誌 A, Vol.J98-A, No.1, pp.29-40, Jan.2015.
- [久行 15] 久行智恵, 三原鉄也, 永森光晴, 杉本重雄: マンガの概念のおよび形態的要素のメタデータに基づくシーン抽出手法の検討, HCG シンポジウム 2015, 2015.
- [山内 14] 山内嶺, 北山大輔: ダイジェスト映像自動生成のための観点の入れ替わりに基づいた特徴的シーン抽出, DEIM Forum 2014, F4-2, 2014.
- [関口 06] 関口一樹, 小杉信, 向井信彦: 映像と音情報を用いた野球中継の自動インデクシング, 電子情報通信学会技術研究報告, 106(398), pp.41-46, 2006-11-24.
- [藤岡 15] 藤岡恭平: マンガの自動要約にむけた要約手法の提案, 早稲田大学, 2015.
- [Chang]Chih-Chung Chang, Chih-Jen Lin. LIBSVM – A Library for Support Vector Machines. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>, (参照 2016/03/25)