

令和 2 年 9 月 17 日現在

機関番号：32689

研究種目：基盤研究(C)（一般）

研究期間：2017～2019

課題番号：17K00511

研究課題名（和文）大規模マンガデータベースのための自動要約生成の研究

研究課題名（英文）Research on Automatic Digest Creation for Large Scale Comic Database

研究代表者

渡辺 裕（WATANABE, HIROSHI）

早稲田大学・理工学術院・教授

研究者番号：10329154

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：ユーザ主導型での高度なマンガ検索を実現することを目的として、機械学習を用いたマンガオブジェクト検出手法及びマンガ要約の自動生成手法の検討を行った。マンガメタデータの対象となるコマ、キャラクタ顔画像、吹き出し、セリフの検出にはFaster R-CNNを用いた方式が有効であることが分かった。画像要約には、マンガの登場人物のキャラクタ顔画像検出した後に、クラスタリングによって分類することが必要でありDBSCANが有効であることが分かった。最後に、自動要約を生成し、主観評価実験により効果を確認した。提案方式は要約の満足度と情報量の点では不十分であるが、理解度については十分な結果が得られた。

研究成果の学術的意義や社会的意義

現在の電子書籍としてのマンガは、その多くが単純な試読ページしか持たず内容把握が容易ではない。しかも多くの電子書籍やマルチメディアデータは、大規模データベースに包含されつつある。そのため、ユーザは所望のマンガに到達することが困難な状況にある。したがって、システムに依存せずにマンガのストーリー把握を可能とするような、エピソード要約や登場キャラクタの提示は、ユーザにとって検索のための重要な情報となる。本研究の成果により、大規模マンガデータから特定のエピソードを検索することが容易になり、自動生成された要約版は電子書籍ユーザにとっての利便性を高める。

研究成果の概要（英文）：In this study, we developed a method for detecting manga objects and automatically generating manga summaries using deep learning techniques to realize user-driven advanced manga search. We found that our method using Faster R-CNN provides the highest detection accuracy for detecting frames, character faces, speech balloons and texts of metadata. For image summarization, it was found that DBSCAN was effective in clustering the images of the character faces after detection.

Finally, automatic summaries were generated and their effectiveness was confirmed by subjective evaluation experiments. The proposed method was inadequate in terms of satisfaction and information content of the summaries, but the results were satisfactory in terms of comprehension.

研究分野：マルチメディア・データベース

キーワード：マンガ 自動要約 機械学習 顔画像検出 キャラクター検出

様式 C-19、F-19-1、Z-19 (共通)

1. 研究開始当初の背景

マンガは内閣官房知的財産戦略本部において、世界的に競争力のある重要なコンテンツであると認識されている。また、アニメの原作供給源としても重要なコンテンツであり、電子書籍の普及と共に、海外に対してもより一層の発信が望まれている。現在の電子書籍としてのマンガは、その多くが単純な試読ページしか持たず内容把握が容易ではない。しかも多くの電子書籍やマルチメディアデータは、大規模データベースに包含されつつある。そのため、ユーザは所望のマンガに到達することが困難な状況にある。したがって、システムに依存せずにマンガのストーリー把握を可能とするような、エピソード要約や登場キャラクタの提示は、ユーザにとって検索のための重要な情報となる。

機械学習に基づいた物体検出アルゴリズムとして 2015 年に発表された Faster R-CNN は候補領域抽出と物体検出を同時に行う手法であり、非常に高速でほぼリアルタイムで物体検出が可能となっている。複数の物体が同時検出可能であるため、マンガメタデータの対象となるコマ、キャラクタ顔画像、吹き出し、セリフなどのオブジェクト検出に適用できる。この時点でのマンガオブジェクト検出手法は、オブジェクトごとに特化された画像特徴量が用いられており、統合的に検出できていなかった。また、マンガは線画やテキストチャから構成されており、通常の一般物体認識のように前後位置が明確ではないため、オブジェクトの重なった画像に対して検出できないという問題があった。

2. 研究の目的

本研究では、ユーザ主導型での高度なマンガ検索を実現することを目的として、ディープラーニングの技術を用いたマンガオブジェクト検出手法及びマンガ要約の自動生成手法の検討を行った。大規模マンガデータから特定のエピソードを検索するためには、要約版が有用である。その生成にはメタデータが必要となる。メタデータに対応するオブジェクトには、コマ割り情報、キャラクタ顔画像、吹き出し、セリフ、背景などがある。本課題では、1)最新のディープラーニング手法である Faster R-CNN を用いてオブジェクトを統合的に高速検出する手法の検討、2)抽出されたメタデータからシーンの重要度を推定し、自動的にマンガ要約を生成する手法の検討、というアプローチを取ることにした。最終的に得られたマンガ自動要約に対して主観評価実験を実施した。

3. 研究の方法

まず、オブジェクト検出のための Faster R-CNN の構成について検討した。マンガ画像に含まれる複数のオブジェクト検出を目的として、Faster R-CNN の学習モデルを用いている。教師データの準備として、多数のマンガデータベースのページに含まれる、コマ、キャラクタ顔画像、吹き出し、セリフなどの位置・サイズ情報を正例データとして登録する。次に、メタデータ検出の高速化・一括化を目指す。要約生成の自動化に向けて、ページ、コマ、キャラクタ顔画像、吹き出し、セリフなどのマンガオブジェクトを関連付けされたメタデータとして登録する。

メタデータは、オブジェクトの種類、位置・サイズ、オブジェクトの包含関係などから構成される。これらのオブジェクトの中で、特にキャラクタ顔画像オブジェクトは、複数の登場キャラクタへの分離が必要となる。エピソード自動要約の対象とするマンガに対して、得られたキャラクタ顔画像に対して、Bug of Visual Words として求められる画像特徴量をクラスタリングすることで、複数の登場キャラクタを抽出する。登場キャラクタを分析することで、メインキャラクタ及び複数のサブキャラクタへの分類が可能となる。また関連するコマのサイズ、吹き出しの大きさ、セリフの多さなどが、重要ページ抽出のためのメタデータとなる。

さらに、マンガの自動要約生成に、ロールコミュニティを適用する。マンガのストーリー展開において、同一シーンに登場するキャラクタの集合を「ロールコミュニティ」として定義し、ロールコミュニティの関連性から重要度を推定する手法により要約生成が可能である。得られた要約を用いることで、大規模マンガデータベースに対するコンテンツアクセスがより容易になると考えられる。最後に主観評価実験を行い提案手法の有効性を示す。

4. 研究成果

多様な形状を持つマンガオブジェクトの検出においては、ディープラーニングの利用が有効と考えられる。しかし、従来の物体検出手法は一般物体検出を目的として設計されていることから、マンガに対する有効性は未知である。したがって、従来手法のマンガオブジェクトに対する検出精度を比較することで、評価を行った。マンガオブジェクトの検出には、マンガ画像に対応した多数のアノテーションが必要となる。そこで、東京大学 相澤・山崎研究室より公開されている研究用マンガデータベースである Manga109 の画像に対して独自にアノテーションの付与作業を行うことでデータセットを構築した。

マンガからシーン単位の情報を取得するには、検出されたキャラクタ顔画像に対してキャラクタごとに分類することで、ラベル付けを行う必要がある。このとき、対象のマンガ作品に関する事前知識がない場合は、登場するキャラクタの総数が分からないという問題がある。これを解決するため、クラスタの数を自動的に決定することが可能な顔画像の分類について検討した。

2017 年度の研究成果として以下の 3 点が挙げられる。

(1) CNN を用いた物体検出手法のうち、Fast R-CNN、Faster R-CNN、SSD の 3 種類につい

てマンガオブジェクト検出に対する有効性を検証した。実験では、それぞれの手法に基づきコマ割り・フキダシ・キャラクタ顔画像・文字列を個別に学習した検出器を作成し、各オブジェクトに対する検出精度を比較した。その結果、4種類のオブジェクトに対する検出精度はFaster R-CNNが最も高くなった(表1, 図1)。

表1 各マンガオブジェクトに対する検出精度の比較 (Average Precision)

	Fast R-CNN	Faster R-CNN	SSD
コマ割り	0.959	0.953	0.897
フキダシ	0.969	0.961	0.907
キャラクタ顔画像	0.810	0.816	0.765
文字列	0.740	0.898	0.866
平均	0.870	0.910	0.859

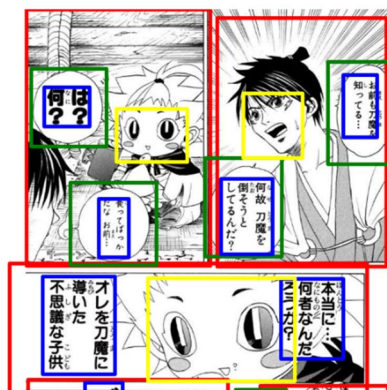


図1 Faster R-CNNによるマンガオブジェクト検出例 (©小林ゆき)

- (2) クラスタの数を自動的に決定するクラスタリング手法である x-means 法のキャラクタ顔画像分類への適用について検討した。実験では、通常の x-means 法では顔画像のクラスタの分割処理が進まないという問題が明らかになったが、分割停止基準を求めるパラメタを調整することでこの問題を解決できることを確認した。
- (3) キャラクタ顔画像のクラスタリングにおける画像特徴量の影響について調査を行った。その結果、CNN の中間出力を特徴量として使用することで、顔特徴やテキストの類似した画像を同一のクラスタとして抽出できることが分かった。

2018年度は、キャラクタ顔画像の分類精度の向上を目標として、画像特徴量及びクラスタリング手法の改良について検討を行った。画像特徴量について、顔画像の分類には同一キャラクタ間で類似度が高くなる特徴量が必要となる。したがって、キャラクタ全般に対して汎化性能を持つ特徴を取得することを目的として、多数のキャラクタ顔画像でファインチューニングを行った CNN を特徴抽出器に使用することを提案した。また、クラスタリングしやすい特徴量を得るために、特徴量の次元削減について検討を行った。

クラスタリング手法について、従来手法はノイズデータの混じった集合に対して正確な分類を行うことが難しいという問題がある。これを解決するため、ノイズに頑健な手法である DBSCAN の適用を提案した。ただし、DBSCAN を使用するにはクラスタの密度の閾値を示すパラメタを設定する必要がある。そこで、最適なパラメタの設定方法について同時に検討を行った。

2018年度の研究成果として以下の3点が挙げられる。

- (4) K-means 法を用いてキャラクタ顔画像をクラスタリングした場合において、顔画像の背景領域の除去が分類精度に及ぼす影響を調査した。実験では、背景を除去した顔画像の集合が一つのクラスタとして抽出されることが原因で全体の分類精度が低下するという結果が得られた。このことから、クラスタリング精度を向上するために、分類するクラスタの数を正確に設定する必要があることが分かった。
- (5) 密度ベースのクラスタリング手法である DBSCAN を用いたキャラクタ顔画像の分類を検討した。実験では、より多くのクラスタを高い分類精度で取得できるかという観点から評価した。その結果、DBSCAN のパラメタを手動で最適な値に設定した場合において、平均で 7.98 個のクラスタを 49.9% の F 値で抽出できることを確認した。また、DBSCAN のパラメタ設定の自動化について検討し、平均で 5.96 個のクラスタを 50.1% の F 値で取得することに成功した。
- (6) キャラクタ顔画像でファインチューニングした CNN を特徴抽出器として使用した顔画像のクラスタリングを検討した。実験では、特徴量を UMAP で 2次元に圧縮した場合において、NMI において最大で 49.5% の精度を示すことが分かった。また、DBSCAN のパラメタ設定を自動化した場合に、NMI において 45.7% の精度で分類を行うことに成功した。(図2)

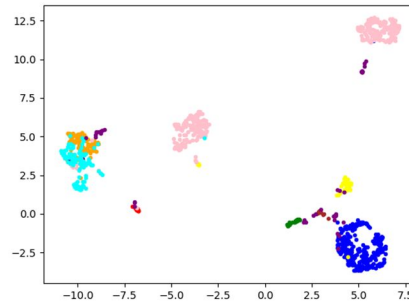


図2 マンガ1作品の登場キャラクターの特徴量の可視化（各点の色がキャラクターに対応）

2019年度は、キャラクター顔画像分類の改善及びマンガ要約手法について検討した。前年度の研究で提案したDBSCANによる顔画像の分類は、DBSCANのパラメタ設定の基準が複雑であるため、汎用性に問題が見られた。したがって、DBSCANを拡張した手法の適用について検討することで、パラメタ設定の簡易化を試みた。

また、これまで得られた研究成果を基にマンガの要約手法を確立するため、画像情報を用いたマンガ要約生成手法を提案し、主観的評価による検証を行った。評価実験では大学生13名を研究補助者として雇用し、要約評価アンケートに回答を得て定量的評価を行った。

2019年度の研究成果として以下の3点が挙げられる。

- (7) キャラクター顔画像のクラスタリングの高精度化を目指して、自己教師あり学習手法であるDeepClusterを組み合わせたCNNの学習について検討した。その結果、DeepClusterを使用することで顔画像の分類精度は低下することが分かった。
- (8) キャラクター顔画像クラスタリングの高性能化を目的として、密度ベースのクラスタリング手法であるOPTICS及びHDBSCANの適用を検討した。DBSCANとOPTICSの比較では、それぞれAMIが最大値をとるようにパラメタを設定した場合において、DBSCANは67.1%、OPTICSは58.7%の精度を示した（表2）。

表2 AMIが最大値をとる場合のDBSCANとOPTICSの比較

	DBSCAN	OPTICS
V-measure	0.685	0.606
ARI	0.610	0.432
AMI	0.671	0.587

- (9) 画像情報を用いたマンガ要約の生成について、主観的評価により有効性の検証を行った。従来研究を基に、「ランダムにページを抽出した要約(要約A)」、「大きなコマを持つページとその前後ページを抜き出した要約(要約B)」、「キャラクターの初登場シーンに着目した要約(要約C)」、「ロールコミュニティモデルを適用した要約(要約E)」の4手法に対してアンケートによるスコア付与を行った。実験結果より、総合的な評価ではランダムに生成した要約に対してその他の手法の有用性を確認することはできなかった。しかし、作品ごとの評価結果では、作品の特色ごとに評価の高い要約が異なるという傾向が分かった(図3)。

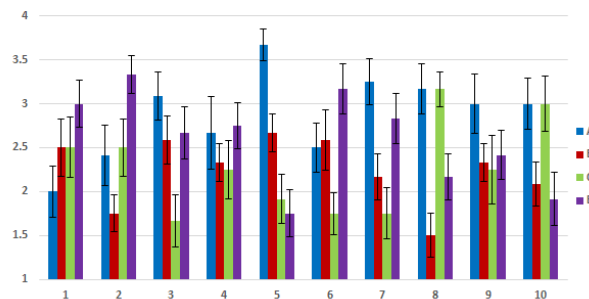


図3 各作品に対する要約A,B,C,Eの要約の質の評価結果（横軸はマンガ作品のid、縦軸はスコアを示す）

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 4件/うち国際共著 0件/うちオープンアクセス 0件）

1. 著者名 Hideaki Yanagisawa, Kengo Kyogoku, Ravi Jain, and Hiroshi Watanabe	4. 巻 20
2. 論文標題 Automatic Classification of Manga Characters using Density-Based Clustering	5. 発行年 2020年
3. 雑誌名 International Workshop on Advanced Image Technology (IWAIT 2020)	6. 最初と最後の頁 1-6
掲載論文のDOI (デジタルオブジェクト識別子) https://doi.org/10.1117/12.2566845	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Hideaki Yanagisawa, Takuro Yamashita, and Watanabe Hiroshi	4. 巻 110491M
2. 論文標題 Manga character clustering with DBSCAN using fine-tuned CNN model	5. 発行年 2019年
3. 雑誌名 Proc. SPIE 11049, International Workshop on Advanced Image Technology (IWAIT) 2019	6. 最初と最後の頁 1-7
掲載論文のDOI (デジタルオブジェクト識別子) https://doi.org/10.1117/12.2521116	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 柳澤 秀彰, 山下 拓朗, 渡辺 裕	4. 巻 73-1
2. 論文標題 主要キャラクタの抽出を目的とした漫画キャラクタ画像のクラスタリング	5. 発行年 2019年
3. 雑誌名 映像情報メディア学会誌	6. 最初と最後の頁 199-204
掲載論文のDOI (デジタルオブジェクト識別子) https://doi.org/10.3169/itej.73.199	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Hideaki Yanagisawa, Takuro Yamashita and Hiroshi Watanabe	4. 巻 16
2. 論文標題 A Study on Object Detection Method from Manga Images Using CNN	5. 発行年 2018年
3. 雑誌名 International Workshop on Advanced Image Technology (IWAIT2018)	6. 最初と最後の頁 1-4
掲載論文のDOI (デジタルオブジェクト識別子) https://doi.org/10.1109/IWAIT.2018.8369633	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計9件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 柳澤秀彰, 京極健悟, Ravi Jain, 渡辺裕
2. 発表標題 密度ベースクラスタリングを用いた漫画キャラクタ顔画像の分類に関する一検討
3. 学会等名 2019年画像符号化シンポジウム・2019年映像メディア処理シンポジウム (PCSJ/IMPS2019)
4. 発表年 2019年

1. 発表者名 柳澤秀彰, 京極健悟, Ravi Jain, 渡辺裕
2. 発表標題 漫画キャラクタ顔画像のクラスタリングにおける特徴表現の改良
3. 学会等名 FIT2019 (第18回科学技術フォーラム)
4. 発表年 2019年

1. 発表者名 柳澤 秀彰, 山下 拓朗, 渡辺 裕
2. 発表標題 CNNを用いた漫画キャラクタ顔画像クラスタリングシステムの改良に関する位置検討
3. 学会等名 映像情報メディア学会年次大会
4. 発表年 2018年

1. 発表者名 山下達郎, 柳澤秀彰, 渡辺裕
2. 発表標題 深層学習を用いたマンガキャラクタの検出における顔変形の影響評価
3. 学会等名 情報処理学会全国大会
4. 発表年 2018年

1. 発表者名 柳澤秀彰、山下達郎、渡辺裕
2. 発表標題 マンガキャラクター顔画像クラスタリングの改良のおける一検討
3. 学会等名 映像情報メディア学会冬季大会
4. 発表年 2017年

1. 発表者名 柳澤秀彰、渡辺裕
2. 発表標題 CNNを用いたマンガオブジェクト検出手法の比較
3. 学会等名 2017年画像符号化シンポジウム・2017年映像メディア処理シンポジウム (PCSJ/IMPS2017)
4. 発表年 2017年

1. 発表者名 山下達郎、柳澤秀彰、渡辺裕
2. 発表標題 深層学習福笑い
3. 学会等名 2017年画像符号化シンポジウム・2017年映像メディア処理シンポジウム (PCSJ/IMPS2017)
4. 発表年 2017年

1. 発表者名 柳澤秀彰、山下達郎、渡辺裕
2. 発表標題 畳込みニューラルネットによるマンガオブジェクト認識メカニズムの一検討
3. 学会等名 電子情報通信学会, パターン認識・メディア理解研究会
4. 発表年 2017年

1. 発表者名 柳澤秀彰、渡辺裕
2. 発表標題 Deep Learning特徴量を用いたマンガキャラクター顔画像の分類
3. 学会等名 FIT2017(第16回科学技術フォーラム)
4. 発表年 2017年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

ADVANCED MULTIMEDIA SYSTEMS LAB. https://www.ams.giti.waseda.ac.jp/publication/

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	柳澤 秀彰 (YANAGISAWA HIDEAKI) (60801280)	早稲田大学・理工学術院・助教 (32689)	令和2年3月退職