

マンガキャラクター顔画像クラスタリングの改良における一検討

A Study on Improvement of Manga Character Face Image Clustering

柳澤 秀彰 山下 拓朗 渡辺 裕

Hideaki YANAGISAWA Takuro YAMASHITA and Hiroshi WATANABE

早稲田大学大学院 基幹理工学研究科 情報理工・情報通信専攻

Waseda University Graduate School of Fundamental Science and Engineering

Abstract Automatic recognition and understanding for manga content has been studied for the purpose of new e-comic services. In the previous research, we proposed character face clustering method using feature vector output from hidden layers in CNN. In this paper, we investigate the influence by removal of background areas.

1. はじめに

電子コミックサービスに有用なメタデータの抽出を目的として、マンガ画像の内容を認識する技術が研究されている。中でも登場キャラクターはマンガの内容を理解する上で重要な要素である。従来研究において、我々はキャラクター顔画像から主要キャラクターを抽出することを目的として、キャラクター顔画像から計算した画像特徴量を X-means 法でクラスタリングすることによる画像分類手法を提案した[1]。しかし、提案手法の問題点として、入力画像の背景がクラスタリングに影響を及ぼすといった点が明らかになった。本稿ではクラスタリング精度の向上を目的として、入力画像の背景削除による影響について検討を行う。

2. X-means 法

X-means 法は K-means 法の逐次繰り返しとベイズ情報量基準(BIC)による分割停止基準を用いて最適なクラスタ数を決定する手法である[2]。BIC の値は以下の式(1)より求められる。

$$BIC = -2 \ln L + k \cdot \ln(n) \quad (1)$$

式(1)において、L はモデルにおける尤度関数の最大値、k はモデルのパラメータの個数、n は標本のサイズを表す。また第 1 項はモデルへの当てはまりの良さを、第 2 項はモデルの複雑さに対するペナルティを示す。X-means 法によるクラスタリング手順は次のようになる。1) はじめに(k=2)として k-means の 2 クラス分類を行い、入力データを 2 つのクラスタに分割する。2) 生成された各クラスタに対して 2-means を行い、分割後のクラスタの BIC が分割前よりも小さいとき分割は適切であるとして採用する。3) 新たに生成されたクラスタに対して 2) の処理を繰り返し、全てのクラスタについて分割後の BIC が分割前より大きくなる状態となったとき、適切なクラスタ数に分割が行われたと見なす。

3. クラスタリング手法

既存手法[1]に従い、SURF 特徴量および CNN の出力をそれぞれ X-means 法によってクラスタリングする。それぞれの手法の詳細を以下に示す。

3.1. SURF 特徴量を使用するモデル

入力画像から SURF 特徴量を計算し、100 次元の Bag-

©長谷川 裕一

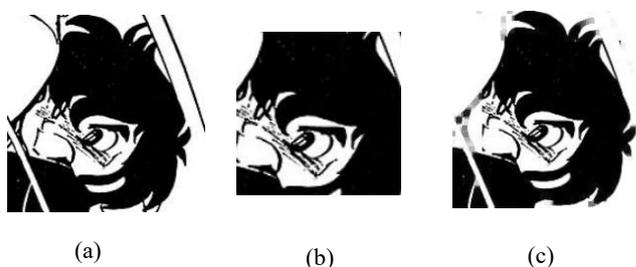


図 1 実験に使用する画像例

(a)元画像 (b)周辺領域を削除した画像 (c)周
辺領域をクロージング処理した画像

of-Visual-Words(BoVW)に変換した特徴ベクトルを X-means に入力する。文献[3]より通常の X-means のパラメータでは分割されるクラスタの数が想定より少なくなるため、本研究では X-means の分割停止基準であるベイズ情報量基準(BIC)の値を以下のように設定する。

$$BIC = -2 \ln L + k \cdot \ln(n) \cdot 0.3 \quad (2)$$

3.2. CNN の出力を使用するモデル

CNN モデルとして AlexNet[4]の一部パラメータを変更したモデルである CaffeNet を使用し、入力重みには ImageNet[5]について学習したパラメータを使用する。fc6 層の出力となる 4096 次元の特徴ベクトルを取り出し、主成分分析によって 100 次元に圧縮したベクトルを X-means に入力する。

4. 実験

研究用マンガ画像データベースの Manga109[6]に公開されているマンガ画像より、作者の異なるマンガ 3 作品 (作品 A: 「BEMADER・P」、作品 B: 「ぶらり鉄扇捕物帳」、作品 C: 「爆裂! かんふー娘」) を使用する。各作品についてキャラクター顔画像 300 枚ずつを抽出し、画像サイズを(width, height)=(200, 200)に正規化したデータセットを作成した。

本研究では背景領域を削除するために、通常の画像に加えて、画像の周囲 10~40 ピクセルを消去した画像

表 1 SURF 特徴量を用いたクラスタリング結果
(その 1)

	元画像	Cut10	Cut20	Cut30	Cut40
作品 A	0.625	0.643	0.519	0.549	0.511
作品 B	0.620	0.777	0.690	0.537	0.506
作品 C	0.650	0.658	0.662	0.634	0.620

表 2 SURF 特徴量を用いたクラスタリング結果
(その 2)

	Closing 10	Closing 20	Closing 30	Closing 40
作品 A	0.576	0.632	0.621	0.566
作品 B	0.635	0.669	0.761	0.710
作品 C	0.754	0.732	0.726	0.786

表 3 CNN の出力を用いたクラスタリング結果
(その 1)

	元画像	Cut10	Cut20	Cut30	Cut40
作品 A	0.627	0.637	0.602	0.609	0.582
作品 B	0.575	0.648	0.624	0.599	0.629
作品 C	0.745	0.737	0.728	0.726	0.689

表 4 CNN の出力を用いたクラスタリング結果
(その 2)

	Closing 10	Closing 20	Closing 30	Closing 40
作品 A	0.624	0.603	0.626	0.615
作品 B	0.632	0.640	0.637	0.610
作品 C	0.737	0.749	0.742	0.744

(cut10~40)および、周囲 10~40 ピクセルに対してクロージング処理を行った画像(closing10~40)を作成した。実験に使用する画像の例を図 1 に示す。SURF 特徴量を使用するモデルと CNN の出力を使用するモデルについてそれぞれ顔画像のクラスタリングを行い、得られたクラスタについて純度(purity)の平均値を比較した。実験結果を表 1 ~ 表 4 に示す。

5. 考察

表 1 および表 2 より、SURF 特徴量を使用したモデルについて、周辺 10 ピクセルを削除した画像に対しては 3 作品ともに元画像からの純度の向上が見られ、それ以上の領域を削除した場合には純度が低下していくことが確認できた。また、クロージング処理を行った場合には作品 B, C に対して純度の向上が見られたが、作品 A では 20 ピクセルの場合を除いて純度は低下する傾向が見られた。

一方で表 3 および表 4 より、CNN の出力を使用したモデルについて、周辺 10 ピクセルを削除した場合において作品 A, B での純度の向上が確認でき、それ以上削除した場合には純度の低下が見られた。また、クロージング処理を行った画像では、作品 B のみ全ての結果において元画像からの純度の向上が見られ、それ以外では、20 ピクセルの場合においてのみ作品 C の精度向上が確認できた。

これらの結果から、周辺領域を削除およびクロージ

ング処理を行うことによって、クラスタリング精度を向上させることは可能だが、その影響はマンガ作品によって異なることが分かった。背景領域を削除する方法では、10 ピクセルを削除することによって背景領域を適切に消去できたためクラスタリング精度の向上に繋がったが、それ以外の場合にはキャラクターの顔特徴に影響を及ぼすため、純度が低下したと考察できる。また、周辺領域に対するクロージング処理は作品 B に対しては有効だが、作品 A ではあまり有効性が見られなかった。この理由として、作品の画風によって背景の傾向が異なることから、クロージング処理によるノイズ除去の有効性が作品によって異なる結果となったと考察できる。

6. まとめ

本稿では、背景領域を削除することによるキャラクター顔画像のクラスタリングの影響について検討を行った。実験結果より、キャラクター顔画像の周辺領域について、削除またはクロージング処理を行うことによってクラスタリング精度の向上が可能であることが確認できたが、どの画像処理が有効であるかはマンガ作品によって異なることが分かった。今後の研究では、より多くのマンガ画像を対象として実験を行うことによって、マンガ作品の画像特徴と背景削除処理の関係について求める。

7. 謝辞

本研究は JSPS 科研費 17K00511 の助成を受けたものである。

文 献

- [1] 柳澤秀彰, 渡辺裕: “Deep Learning 特徴量を用いたマンガキャラクター顔画像の分類”, FIT2017(第 16 回科学技術フォーラム), H-001, (Sep. 2017)
- [2] D. Pelleg, A. Moore: “X-means: Extending K-means with efficient estimation of the number of clusters” In Proceedings of the 17th International Conf. on Machine Learning, pp.727-734, (2000)
- [3] 柳澤秀彰, 渡辺裕: “X-means 法を用いたマンガキャラクターの自動分類に関する検討”, 電子情報通信学会総合大会, D-12-40, (2017)
- [4] A. Krizhevsky, I. Sutskever, G. Hinton: “Imagenet classification with deep convolutional neural networks” Advances in Neural Information Processing Systems, (2012)
- [5] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei: “ImageNet: A Large-Scale Hierarchical Image Database”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (2009)
- [6] Y.Matsui, K.Ito, Y. Aramaki, T.Yamasaki, K. Aizawa: “Sketch-based Manga Retrieval using Manga109 Dataset”, arXiv:1510.04389, (2015)

早稲田大学大学院 基幹理工学研究科

情報理工・情報通信専攻

〒169-0072 東京都新宿区大久保 3-14-9

早大シルマンホール 401

TEL.03-5286-2509 E-mail: bule-cosmo@ruri.waseda.jp