

漫画キャラクター顔画像のクラスタリングにおける特徴表現の改良 Improvement of Feature Representation in Clustering of Manga Character Faces

柳澤 秀彰[†] 京極 健悟[‡] Ravi Jain[‡] 渡辺 裕[†]
Hideaki Yanagisawa Kengo Kyogoku Ravi Jain Hiroshi Watanabe

1. はじめに

現在、電子書籍市場の売上の80%以上を電子コミックが占めているとされる。この背景から、電子コミックの利便性を向上させる技術として、漫画の内容をクエリとした高度検索や、漫画画像を入力としたストーリー要約の生成などのシステムが提案されている。

漫画のストーリーを理解するには、漫画内に登場するキャラクターの認識が必要となる。未知の漫画作品を対象としたキャラクターの認識を行うための手法として、キャラクター顔画像から抽出した画像特徴量をクラスタリングすることによる分類が提案されている。このとき、漫画に対して汎化能力を持った画像特徴を得るために、キャラクター顔画像のクラス分類問題で *fine-tuning* を行った CNN が特徴抽出器として利用されている。しかし、CNN の *fine-tuning* にはクラスラベルを付与した漫画画像データが必要となるため、データセットの拡張が困難であるという問題が存在する。また、漫画のキャラクターは作品ごとの形状変化が大きいいため、クラス分類だけでは汎化性能の高い特徴を学習することが困難であると考えられる。

このような問題に対して、本研究では教師データの存在しない入力に対して自己教師あり学習で特徴表現を取得する手法である DeepCluster のキャラクター顔画像クラスタリングへの適用を提案する。実験では、従来の *fine-tuning* および、DeepCluster によって学習したモデル、さらに少数のデータセットで *fine-tuning* した重みを使い DeepCluster の学習を行ったモデルの3種類について HDBSCAN によるクラスタリングを適用し、クラスタリング精度を比較した。

2. 関連研究

未知の漫画作品を対象としてキャラクターを認識する研究について、Narita らは漫画画像から特定のキャラクターを検索するシステムを提案している[1]。この手法では、キャラクター顔画像で *fine-tuning* した CNN を特徴抽出器として入力画像と漫画画像の類似度を比較している。また、坪田らは CNN の深層特徴をベースとした *k-means* 法によるキャラクター顔画像のクラスタリングを提案した[2]。この研究では、対象の漫画作品について深層距離学習を行うことで、個々の作品に対するクラスタリング精度を向上させることに成功している。我々は、登場キャラクターのクラス数が未知の状態におけるキャラクターの認識を目的として、次元削減した CNN 特徴量を入力として、DBSCAN でクラスタリングする手法を提案した[3]。

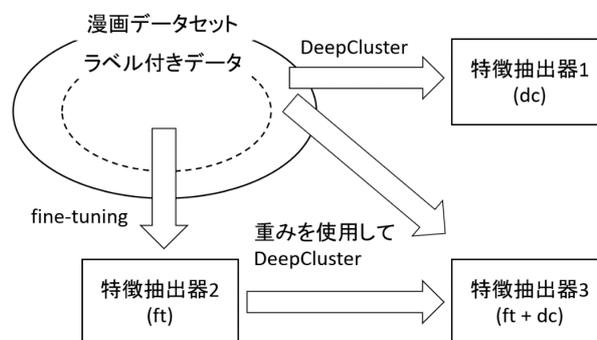


図1 提案手法の概要

3. 提案手法

ラベルが存在しないキャラクター顔画像から汎化能力のある特徴表現を取得することを目的として、DeepCluster によるキャラクター顔特徴の学習を提案する。また、本研究では登場キャラクターの数が未知であることを想定し、特徴量のクラスタリング手法に HDBSCAN を使用する。

3.1 DeepCluster

Carlson らは、ニューラルネットワークを使った自己教師あり学習によってクラスタリングを行うアーキテクチャとして DeepCluster を提案した[4]。この手法は、入力データに対する疑似ラベルの付与と、疑似ラベルを用いた分類器の学習の二つの段階からなる。まず、CNN に入力された画像データの特徴計算を行い、最終層の一つ前の層の出力を特徴量として抽出する。次に、特徴量を *k-means* でクラスタリングし、生成されたクラスターの番号を疑似ラベルとする。さらに、再度 CNN の特徴計算を行い、最終層から得られたクラス番号と疑似ラベルから *cross entropy loss* を求め、逆伝搬により重みを更新する。この処理を各 *epoch* で繰り返えし、教師なしデータから学習済みモデルを生成する。

本研究では、大規模なキャラクター顔画像データの中に、ラベルが付与された画像データが一部だけ存在する状況における DeepCluster による特徴表現の学習を提案する。提案手法の概要を図1に示す。データセット全体から DeepCluster の学習を行ったモデルを「dc」、ラベル付きデータを入力として *fine-tuning* を行ったモデルを「ft」、ft の重みを初期値として、データセット全体から DeepCluster の学習を行ったモデルを「ft + dc」とする。

3.2 HDBSCAN

データの密度に基づいたクラスタリング手法として、Ester らは DBSCAN (Different Based Spatial Clustering of Applications with Noise) を提案した[5]。この手法は、各点 p について距離 ϵ の範囲内に存在するデータを数え、範囲内

[†] 早稲田大学 基幹理工学部 情報通信学科

Department of Communications and Engineering, Waseda University

[‡] 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

Department of Computer Science and Communications and

Communications Engineering, Waseda University

にデータ点が minPts 個以上存在する点をコア点とおく。そして、コア点と隣接したデータの集合を一つのクラスタとして抽出する。DBSCAN では前もってクラスタ数を設定する必要がなく、また外れ値の影響を受けにくいという利点を持つ。一方、クラスタリング結果がパラメータ ϵ の値に大きく依存するという問題がある。

HDBSCAN は DBSCAN の拡張として、パラメータ ϵ の代わりに階層クラスタリングによって有効なクラスタリング結果を求める手法である[6]。まず、各コア点の最近傍距離を求め、最小全体木を構築する。次に、全体木をエッジの距離が大きい順に並べ替えることで、距離の閾値によって隣接したクラスタがマージされる関係を表した樹形図を求める。さらに、樹形図から最小クラスタサイズよりも少ない枝を除外して要約する。最後に、樹形図の中で距離の変化に対してマージされずに存在する期間の長いクラスタが安定したクラスタであるとして抽出する。

4. 実験

3.1 項で述べた 3 種類の特徴抽出器について、未知の漫画画像に対するクラスタリング精度を比較する。本実験では、Manga109[7][8]で公開されている漫画画像を使用し、キャラクター顔領域を切り出したデータセットを作成する。CNN の学習に使用するデータセットは、作者の異なる漫画 83 作品から 10 回以上登場しているキャラクターを対象に顔画像を抽出した 1222 クラスの画像からなり、訓練画像を 77076 枚、テスト画像を 6110 枚とする。また、テストセットのうち 20 作品からなる 257 クラスの画像をラベル付きデータとして扱い、訓練画像 16454 枚、テスト画像 1285 枚で構成する。クラスタリングの評価に使用するデータセットは、学習データとは異なる 11 作品から作品別に顔画像を抽出して作成する。このとき、作品内で登場回数が 10 枚以上のキャラクターは個別のクラスを割り当て、10 回未満のキャラクターは「その他」として 1 つのクラスと見なす。

CNN の構成は VGG16 を使用する。DeepCluster の設定は dc, ft + dc とともに、学習率を 10^{-4} 、k-means のクラスタ数を 3000、epoch 数を 200 とする。fine-tuning の設定は学習率を 10^{-4} 、epoch 数を 200 とする。また、HDBSCAN の設定は最小クラスタサイズを 10、コア点の最小サンプル数を 1 とする。クラスタリングに使用する特徴量は VGG16 の max pooling 層の直前の各畳み込み層および、最終層を除く fc 層の出力を使用し、UMAP で 32 次元に次元削減する。クラスタリングの評価基準には AMI を使用する。

3 つのモデルによるクラスタリング結果を表 1 に示す。この結果より、3 種類のモデルともに 1 番目の fc 層の出力を使用した時に AMI の値が最も高くなることが分かった。また、小規模データセットで fine-tuning したモデルよりも DeepCluster で学習したモデルの方が高い精度が得られた。fc1 層の出力を使用した場合における作品別のクラスタリング結果を図 2 に示す。この図より、dc と ft は作品によって有効性の変動が大きい一方で、ft + dc ではどの作品に対しても 2 位以上の精度を示していることが分かった。このことから、教師あり学習と DeepCluster を組み合わせることによって両手法の欠点を補うことができると考えられる。

5. まとめ

本研究では、一部の画像のみにラベルが付与された漫画画像データセットを対象として DeepCluster を用いた特徴学

表 1: クラスタリング結果の比較 (AMI)

	conv1	conv2	conv3	conv4	conv5	fc1	fc2
dc	0.168	0.255	0.330	0.338	0.382	0.418	0.400
ft	0.251	0.238	0.305	0.347	0.311	0.362	0.351
dc+fc	0.278	0.306	0.367	0.405	0.421	0.427	0.375

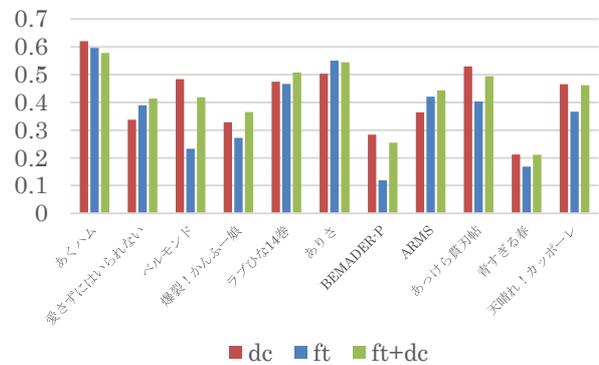


図 2 各作品に対するクラスタリング結果 (AMI)

習による顔画像クラスタリングを提案した。実験結果より、DeepCluster によって従来の fine-tuning よりも汎化性能に優れた特徴抽出器が生成可能であることを示した。また、DeepCluster の初期重みに fine-tuning で得られた重みを使用することで、クラスタリング精度を向上させることに成功した。今後の課題として、特徴量のクラスタリング方法の改良や、より大規模な漫画画像データベースを入力とした場合の有効性の検証が挙げられる。

謝辞

本研究は JSPS 科研費 17K00511 の助成を受けたものである。

参考文献

- [1] R. Narita, K. Tsubota, T. Yamasaki, and K. Aizawa, "Sketch-based Manga Retrieval using Deep Features", 2017 14th IAPR International Conference on Document Analysis and Recognition, pp.49-53 (2017).
- [2] 坪田亘記, 小山徹, 山崎俊彦, 相澤晴彦, "キャラクター顔特徴量の個別漫画への適応手法", 人工知能学会全国大会(32回), 1K2-OS-2b-04 (2018).
- [3] 柳澤秀彰, 山下拓朗, 渡辺裕, "主要キャラクターの抽出を目的とした漫画キャラクター画像のクラスタリング", 映像情報メディア学会オンラインジャーナル, vol.73, No.1, pp.199-204 (2019).
- [4] M. Caron, P. Bojanowski, A. Joulin, M. Douze, "Deep Clustering for Unsupervised Learning of Visual Features", European Conference on Computer Vision, <https://arxiv.org/abs/1807.05520> (2018).
- [5] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise" in Proceeding of 2nd International Conference on Knowledge Discovery and Data Mining, pp.226-231, (1996).
- [6] L. McInnes, J. Healy, S. Astels, "hdbscan: Hierarchical density based clustering", Journal of Open Source Software, vol.2, No.11, pp.205 (2017).
- [7] Y.Matsui, K.Ito, Y.Aramaki, A.Fujimoto, T.Ogawa, T.Yamasaki, K.Aizawa, "Sketch-based Manga Retrieval using Manga109 Dataset", Multimedia Tools and Applications, Springer (2017).
- [8] T.Ogawa, A.Otsubo, R.Narita, Y.Matsui, T.Yamasaki, K.Aizawa, "Object Detection for Comics using Manga109 Annotations", arXiv:1803.08670, (2018).