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

A Study about Classification of Manga Character Faces using Density-Based Clustering

柳澤秀彰† Hideaki Yanagisawa†
京極健吾‡ Kyogoku Kengo‡
Jain Ravi‡ Jain Ravi‡
渡辺裕† Hiroshi Watanabe†

早稲田大学 基幹理工学部 情報通信学科†
Waseda University, Graduate School of Fundamental Science and Engineering, Department of Communications and Computer Engineering†

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, Department of Computer Science and Communications Engineering‡

Abstract: 未知の漫画作品から登場キャラクタの情報を自動的に抽出するためには，ドメイン知識を使わずにキャラクタを認識する技術が必要となる．従来研究では，キャラクタ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で，類似した特徴を持つ画像を同一のキャラクタとして分類する手法を提案している．我々はクラスタリング手法に DBSCAN を利用することで，クラスタ数を指定することなくキャラクタを分類する手法を提案した．本研究では，キャラクタ顔画像のクラスタリングの改良として，DBSCAN を拡張した密度ベースクラスタリングの適用を検討する．また，CNN の学習に深層距離学習を適用することによる性能変化について検討を行った．

1 はじめに

漫画に登場するキャラクタの情報は，ストーリーを理解する上で重要な要素の一つである．未知の漫画作品からキャラクタの情報を抽出するには，正解ラベルやクラスタ数といったドメイン知識を用いずにキャラクタを認識する技術が必要である．本研究は，密度ベースのクラスタリングを用いたキャラクタ顔画像の分類について検討を行う．さらに，深層学習を特徴抽出器の学習に用いる効果について検証する．

2 従来研究

未知の漫画作品を対象としてキャラクタを分類する方法として，坪田らはキャラクタ顔画像で fine-tuning を行った畳み込みニューラルネットワーク(CNN)を特徴抽出器として利用し，出力された特徴ベクトルを k-menas 法でクラスタリングする手法を提案した[1]．この研究では，認識の対象とする漫画作品について CNN の深層距離学習を行うことで分類精度を向上させることができると報告している．しかし，k-means によるクラスタリングは予め生成するクラスタ数を指定する必要があるという問題がある．我々は，クラスタ数を指定せずにキャラクタを分類することを目的として，CNN の出力した画像特徴量を次元圧縮し，距離ベースのクラスタリング手法である DBSCAN によって分類する手法を提案した[2].

3 提案手法

キャラクタ顔画像で fine-tuning した CNN の中間出力を画像特徴量として使用し，これをクラスタリングすることでキャラクタ顔画像の分類を行う．本研究では，クラスタリングに用いる手法として，従来手法の DBSCAN の他に DBSCAN の拡張である OPTIC と HDBSCAN の二つについて比較を行う．また，特徴抽出器の性能向上を目的として，深層距離学習の手法の一つである ArcFace を用いて fine-tuning を行うことによる効果を検証する．

3.1 特徴抽出器

本研究は CNN のモデルに ResNet50 を使用し，全結合層の直前の層の出力を特徴量とする．さらに出力された特徴量を UMAP[3]によって 32 次元に圧縮し，クラスタリングを適用する．CNN の fine-tuning には Manga109[4, 5]データセットの漫画画像を使用する．学習セットには作者の異なる 83 作品から 10 回以上登場するキャラクタについて顔領域を切り出した画像を使用し，これらをキャラクタ別にクラス分けする．顔領域はデータセットのアノテーションで指定された領域を縦横 2 倍ずつ拡張した領域と定め，領域がページ内に収まるようにクリッピングする．これは，顔領域の周辺にも髪型などのキャラクタを表す情報が含まれていると考えられるためである．データセットは 1,222 クラスの画像 77,076 枚から成る．

本研究では，よりクラスタリングに適した特徴表現を得ることを目的として，特徴抽出器の学習に深層距離学習の一つである ArcFace を用いた場合の影響を検討する．ArcFace は通常のクラス分類の学習に式(1)のロス関数を適用することで，クラス内の分散を小さくし，異なるクラス間の分散を大きくする手法である．

$$-\frac{1}{N}\sum_{i=1}^{N}\log\frac{\exp\left(s\cos\left(\theta_{y_i}+m\right)\right)}{\exp\left(s\cos\left(\theta_{y_i}+m\right)\right)+\sum_{k\neq y_i}\exp\left(s\cos\left(\theta_k\right)\right)} \quad (1)$$

式(1)において，y_iは入力データ$\boldsymbol{x}$の正解クラスであり，$m(>0), s(>1)$はそれぞれ重みのパラメタである．ArcFace は特徴ベクトルをL_2正規化し，正解クラスにのみ重みmを付与したコサイン距離を計算する．

3.2 密度ベースクラスタリング

DBSCAN はデータの密度に基づいてクラスタを生成する手法である．DBSCAN の実行には最小クラスタサイズ(MinPts)と到達可能距離(Eps)の 2 つのパラメタが必要となる．半径 Eps 以内に MinPts 個以上のデータが存在する点をコア点と定め，連結したコア点の集合を一つのクラスタとして抽出する．従って，

DBSCAN のクラスタリング結果は Eps の設定によって大きく変動する．[2]で提案した手法では，MinPts を固定値として，入力された特徴ベクトルについて k 近傍距離のグラフを求め，グラフの勾配変化が起きる点を Eps に設定することでクラスタリングの自動化を実現した．一方，パラメタ Eps の設定に依存しないように DBSCAN を拡張した手法として OPTICS と HDBSCAN が提案されている．

OPTICS は空間データから密度ベースのクラスタを求める手法である．まず，データ間の最近傍距離を求め，データセット全体における到達可能性プロットを作成する．次に，特定の閾値を定めることで到達可能性プロットを分断し，MinPts 以上のクラスタサイズを持つ集合を抽出する．この閾値は到達可能性プロットの勾配情報から求めることが可能である．

HDBSCAN はクラスタの安定性に基づきフラットなクラスタを求める手法である．始めに，データ間の最小距離を求め，データ間距離とデータの結びつきの関係を示す樹形図を生成する．次に，樹形図から要素の数が MinPts 個を満たさない枝を削除し，残った枝をそれぞれクラスタとする．最後に，生成されたクラスタC_iの評価指標$S(C_i)$を式(2)から求め，$S(C_i)$の組み合わせが最大となるクラスタ群を抽出する．

$$S(C_i) = \sum_{x_j \in C_i} \left(\lambda_{max}(x_j, C_i) - \lambda_{min}(C_i) \right) \tag{2}$$

ここで，$\lambda_{min}(C_i)$はクラスタC_iが成立する最小の密度の値であり，$\lambda_{max}(x_j, C_i)$はデータx_jがクラスタC_iに含まれる条件下での最大の密度を示す．

4 実験

キャラクタ顔画像で fine-tuning した CNN を特徴抽出器として，DBSCAN, OPTICS, HDBSCAN によるクラスタリング結果を比較する．また特徴抽出器について，通常のクラス分類を学習した CNN と ArcFace で学習した CNN のクラスタリングにおける性能を比較する．

4.1 データセット

評価には Manga109 より作者の異なる 11 作品から切り出したキャラクタ顔画像合計 9,696 枚を使用する．各作品について 10 回以上登場するキャラクタを個別のクラスに割り当て，登場回数が 10 回に満たないキャラクタは「その他」のクラスに割り当てる．

4.2 クラスタリング設定

DBSCAN, OPTICS, HDBSCAN について，MinPts の値は 10 に設定する．また DBSCAN のパラメタ Eps は[2]で提案した手法に基づいて設定する．OPTICS の閾値を決定する最小急勾配 xi の値は 0.30 に定める．

4.3 評価方法

クラスタリングの評価基準には Adjusted Mutual Information (AMI)を用いる．正解ラベルと生成されたクラスタの集合をそれぞれ$U = \{u_1, u_2, \cdots, u_N\}$，$V = \{v_1, v_2, \cdots, v_N\}$としたとき，AMI は相互情報量(MI)を用いて以下の式(3)で求められる．

$$AMI(U,V) = \frac{MI(U,V) - E\left(MI(U,V)\right)}{1/2\{H(U) + H(V)\} - E\{MI(U,V)\}} \tag{3}$$

AMI のスコアは 0.0~1.0 の範囲であり，クラスタリング結果が正解ラベルに近いほど高い値が得られる．なお，本実験ではクラスタリングにおいて外れ値と判定されたデータの集合を 1 つのクラスタと扱う．また，乱数を考慮して 10 回クラスタリングを行った平均値を比較する．

表 1: クラスタリング手法の比較(AMI)

	DBSCAN [2]	OPTICS	HDBSCAN
Original	0.628	0.547	0.658
ArcFace	0.346	0.110	0.338

4.4 実験結果

3 種類のクラスタリング手法および特徴抽出器の学習方法について比較を行った結果を表 1 に示す．表 1 において「Original」は従来のクラス分類を学習した特長抽出器を示す．クラスタリング手法において，従来の DBSCAN と比較すると，OPTICS では AMI が低下した一方で, HDBSCAN では AMI の向上が見られた．この結果より，HDBSCAN を導入することで，手動で有効なパラメタを求める従来手法よりも有意な結果が得られることが分かった．

また，特徴抽出器の比較では ArcFace を適用することで AMI が低下することを確認した．この理由として，ArcFace によって得られた特徴表現は既知のクラスのキャラクタを対象とした分類に特化しているため，学習データと異なる特徴を持った未知のキャラクタに対する特徴記述性能は低下したからであると考察できる．

5 おわりに

本研究では，密度ベースのクラスタリングを用いたキャラクタ顔画像の分類について検討した．実験結果より, クラスタリング手法として HDBSCAN が有効であることを確認した．また，特徴抽出器の fine-tuning に ArcFace を用いることは，未知のキャラクタを対象としたクラスタリングにおいては有効でないという知見が得られた．

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] 坪田亘記，小川徹，山崎俊彦，相澤清晴，“キャラクター顔画像の個別漫画への適応手法”，人工知能学会全国大会(第 32 回)，1K2-OS-2b-04, 2018.

[2] H. Yanagisawa, T. Yamashita and H. Watanabe, “Character Clustering with DBSCAN using Fine-Tuned CNN Model,” Proc of International Workshop on Advanced Image Technology and International Forum o Medical Imaging in Asia, 2019.

[3] L. Mclnnes, J. Healy, “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction”, arXiv:1802.03426, 2018.

[4] 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.

[5] T.Ogawa, A.Otsubo, R.Narita, Y.Matsui, T.Yamasaki, K.Aizawa, “Object Detection for Comics using Manga109 Annotations”, arXiv:1803.08670, 2018.

早稲田大学　基幹理工学部　情報通信学科
〒169–0072 東京都新宿区大久保 3-14-9
シルマンホール 401
Phone: 080–1252–2394