# Automatic Classifiaction of Manga Characters using Density-Based Clustering

Hideaki Yanagisawa, Kengo Kyogoku, Ravi Jain and Hiroshi Watanabe



December 2, 2019

# Automatic Classification of Manga Characters using Density-Based Clustering

Hideaki Yanagisawa*[a], Kengo Kyogoku[b], Jain Ravi[b], Hiroshi Waranabe[a]
[a] Waseda University Graduate School of Fundamental Science and Engineering Department of Communications and Computer Engineering, Shilman Hall 401, 3-14-9 Okubo, Shinjuku, Tokyo, Japan 169-0072; [b] Waseda University Graduate School of Fundamental Science and Engineering Department of Computer Science and Communications Engineering, Shilman Hall 401, 3-14-9 Okubo, Shinjuku, Tokyo, Japan 169-0072

## ABSTRACT

Manga (Japanese comic) is a globally popular content. In recent years, sales of e-comics that converted to electronic data from paper-based manga are increasing because of the widespread use of electronic terminals. Against this background, it has been proposed to improve the accessibility of e-comics by tagging manga images with metadata. In order to allocate metadata more efficiently, technology that automatically extracts elements such as character and speech is required. One way to classify characters is to get image features from the character's faces and cluster them. Previous research has shown that using the intermediate output of CNN which fine-tuned with character face images is effective for character face recognition. We proposed a clustering method using Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to classify character face images without specifying the number of clusters. However, DBSCAN is greatly affected by the hyperparameter. The purpose of this study is to automatically classify character face images without complicated hyperparameter setting. We examine the application of Ordering Points to Identify the Clustering Structure (OPTICS) and Hierarchical DBSCAN (HDBSCAN), which are density-based clustering algorithms that extend DBSCAN. OPTICS is an algorithm for finding clusters in spatial data, and HDBSCAN is an algorithm extracts flat partition from hierarchical cluster data. We also verify the effective CNN model as the feature extractor of face images. Experimental results showed that HDBSCAN is effective for character face image clustering.

**Keywords:** manga, clustering, density-based, DBSCAN, OPTICS, HDBSCAN

## 1. INTRODUCTION

Manga (Japanese comic) is a globally popular content. In recent years, sales of e-comics are increasing because of the widespread use of electronic terminals. In Japan, e-comic sales account for 80% of all e-books market. E-comics currently distributed are composed of simple image files, but new services can be given by tagging metadata to e-comics. For example, a retrieval system for e-comics using the elements of manga such as characters and speeches and automatic generation of manga summaries based on the elements have been proposed. In order to tag metadata efficiently, a technique for automatically extracting manga elements is required. Manga characters are necessary elements to understand manga stories, and are drawn in completely different ways for each manga book. Therefore, in order to analyze unlabeled manga, it is necessary to classify characters without using domain knowledge such as names and numbers of characters. In our previous research, we proposed a method for classifying character face images by using CNN as an image feature extractor and clustering with Density-Based Spatial Clustering of Applications with Noise (DBSCAN). DBSCAN is an algorithm that generates clusters according to the distribution density of input data. This algorithm automatically determines the number of clusters to generate. However, clustering result of DBSCAN is greatly affected by hyperparameter setting. In this research, we study about application of other density-based clustering methods to simplify character face clustering. In addition, we evaluate structure of CNN models used to extract the face features of characters.

*hideaki1@aoni.waseda.jp; phone 81 080 1252-2394

# 2. RELATED RESEARCH

Nagao et al. clustering method of character face images to recognize characters in manga by unsupervised learning. [1] This method is based on the idea that face images of same character have similar image features, and uses Speeded Up Robust Features (SURF) to express the image features of character faces. Tsubota et al. proposed a clustering method that uses the output of CNN fine-tuned by training general character faces as the image features. [2] In this paper, they proposed the adaptation of deep metric learning to image feature training to further improve the recognition accuracy of individual manga. However, since these methods use clustering by k-means, it is necessary to specify the number of clusters to be generated beforehand. Therefore, there is a problem that accurate clustering cannot be performed unless the number of characters that appear is known. To solve this problem, we proposed a clustering of facial image features using DBSCAN. DBSCAN generates clusters based on data density and can automatically determine the number of clusters. [3] Our method succeeded in automatically classifying character face images without domain knowledge.

# 3. PROPOSED METHOD

DBSCAN has the advantage that it automatically determines the number of clusters and is robust against noise, but it has the disadvantage that clustering result greatly depends on the hyperparameter settings. OPTICS and HDBSCAN are methods to detect valid clusters by applying DBSCAN density-based clustering to the hierarchy. In this study, we compare the effectiveness of three methods for clustering of character face images. The basic flowing of character face image clustering is as follows. First, image features are obtained by inputting character face images to fine-tuned CNN. Next, they are classified using following clustering algorithms.

### 3.1 Feature extractor

In this study, the output of the middle layer of a CNN fine-tuned by training general characters' face images is used as the image features. The CNN is pre-trained on ImageNet [4] and is fine-tuned by training classification face images of characters appearing in 83 manga each drawn by different authors. The final full connection layer is removed and an L2 normalization layer is added in the last layer to obtain image features. For training, we use manga images in Manga109 [5, 6] dataset. The training set is created by cutting out the face regions of characters who appear more than 10 times in each book. The face regions are defined as the rectangle regions with twice height and width from the face regions specified by the annotation in the dataset, and clipped to fit within the page region. The training set consists of 77,076 images in total for 1,222 classes. The outputs obtained from the CNN are high-dimensional feature vectors. It is known that clustering high-dimensional data has problems such as increase in calculation time and decrease in clustering accuracy. Therefore, the dimensions of feature vector are reduced by UMAP [7] to reduce the computation cost.

### 3.2 DBSCAN

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a clustering algorithm that extracts clusters of arbitrary shapes. [8] It uses two parameters: distance threshold *Eps* and data count threshold *minPts*. The connection relation of data points is defined from these parameters, and the connected data points are classified into the same cluster. This connection relationship is called directory density-reachable (DDR), and it is said that from data point $x_p$ to $x_q$ is DDR when the following conditions are satisfied.

$$x_q \in N_{Eps}(x_p) \tag{1}$$

$$\left|N_{Eps}(x_p)\right| \geq Minpts \tag{2}$$

When, $N_{Eps}(x_p) = \{x_p \in X \mid D(x_p, x_q) \leq Eps\}$. $D(x_p, x_q)$ means the Euclidean distance between $x_p$ and $x_q$. This relationship is not symmetric. Then, the largest data set that connected with DDR is extracted as a cluster. The algorithm sequentially classifies data points DDR from arbitrary data points into clusters. If no new DDR data point can be found, a new data point is selected, excluding the data points in the generated cluster. DBSCAN can extract clusters with complex shapes, and even in the presence of noise, if *minPts* can be set appropriately, only dense data sets can be extracted as clusters. In this study, the value of the parameter *Eps* is determined based on the conventional method. [3] This method uses an algorithm for extracting *Eps* candidates proposed by Soni et al. [9] First, calculate the influence function consisting of the Euclidean distance between two points $x$ and $y$.

$$INF(x,y) = \sqrt{\sum_{i=1}^{d}(x_i - y_i)^2} \tag{3}$$

Where $d$ means the data dimension. Second, the local density function which is the distance among the point $x$ and k-nearest neighbors (k = *minPts*) is calculated.

$$DEN(x, y_1, \cdots, y_k) = \sum_{i=1}^{k} INF(x, y_i) \tag{4}$$

The local density function is the average as follows.

$$ADEN(x, y_1, \cdots, y_k) = \frac{\sum_{i=1}^{k} INF(x, y_i)}{k} \tag{5}$$

Then, normalize the AGED value for each point and make into a histogram. Next, extract buckets containing k or more elements from the histogram. Finally, an average value of each extracted bucket is obtained as candidate of *Eps* value. We name these $n$ candidates to $c_1, c_2, \cdots, c_i, \cdots c_n$. In the proposed method, the slopes of $c_i$ and $c_{i+1}$ are calculated in ascending order, and the *Eps* value is set to $c_i$ when the slope changes higher than the average slope. However, since the slope between the first pair is often higher than the average, $c_1$ is excluded.

### 3.3 OPTICS

Ordering Points to Identify the Clustering Structure (OPTICS) is an algorithm that find density-based clusters in spatial data proposed by Ankerst et al. [10] Its basic idea is similar to DBSCAN. However, it is improved to support cluster detection from data with different densities. This method sorts the data points according to the distance from the nearest neighbor, and extracts clusters by creating a special kind of dendrogram. The flowing of OPTICS algorithm is shown follow. OPTICS also uses two parameters: *Eps* and *minPts*. A point $x_p$ is a core point if at least *minPts* points are found within the radius *Eps*. Unlike DBSCAN, OPTICS considers more dense points in the cluster, so each point is assigned a core distance that represents the distance to the *minPts*-th closest point. The reachability-distance of another point $x_q$ from a point $x_p$ is either the distance between $x_q$ and $x_p$, or core distance of $x_p$, whichever is bigger. The parameter *Eps* is not mandatory in OPTICS, so it can be simply set to the maximum possible value. By removing this parameter, OPTICS abstracts from DBSCAN. Next, the reachability-plot is used to obtain the cluster hierarchy. This is a two-dimensional plot, where x-axis shows ordering of the points as processed by OPTICS, and y-axis shows the reachable distance. Since the points belonging to clusters have low reachable distances to their nearest neighbors, the clusters are represented as valleys in the reachability plot. The deeper the valley, the denser the cluster. Clusters can be extracted from this plot by using algorithms that detect valleys with steep slopes, knee detection, and local maxima.

### 3.4 HDBSCAN

Hierarchical DBSCAN (HDBSCAN) is a clustering algorithm developed by Campello et al. [11] It extends DBSCAN by converting it into a hierarchical clustering algorithm, and then using a technique to extract a flat clustering based in the stability of clusters. The process of HDBSCAN is as follows. First, remove low-density points that cause noise. According to the parameter "*minSamples*", the kth nearest neighbor distance (k = *minSamples*) is determined as the core distance. Points with a longer core distance are weighted to increase the reachable distance. Second, a minimum spanning whose vertices are data point and edges between two points are equal to the reachable distance is created. By applying thresholds to edges, a hierarchy of components connected at various threshold levels is obtained. Third, nodes with fewer points than *minPts* are deleted to compress the tree. Finally, stable clusters are extracted from the tree. In this process, $\lambda = \frac{1}{distance}$ is used as a measure to consider cluster persistence. For the specified cluster, it is defined that $\lambda_{birth}$ as the value when the cluster is divided into its own cluster, and $\lambda_{death}$ when the cluster is divided into smaller clusters. In turn, for each point $p$ in a cluster, the lambda value that point 'fell out the cluster' is define as $\lambda_p$. This value exists between $\lambda_{birth}$ and $\lambda_{death}$. Now, for each cluster compute the stability as

$$\sum_{p \in cluster}(\lambda_p - \lambda_{birth}) \tag{6}$$

The tree is searched from the end node to the root node, and the combination of the clusters with the highest stability is extracted.

## 4. EXPERIMENT

In this experiment, we compare the performance of density-based clustering in the classification of manga characters. The architectures of the CNN model used for character face image feature extraction are VGG16,[12] ResNet50,[13] and ResNet101.[13] Apply fine-tuning to each model and make feature extractor as described in Section 3.1. For evaluation, we use 11 manga books drawn by different authors from the manga in training set. All character face images for evaluation are cut out same as training images. For each comic book, create a file that places the face images of characters appearing 10 times or more in separate classes and places other characters in the "Other" class. Table 1 shows the contents of the evaluation set.

Table 1. Manga titles, number of classes and number of character face images included in the evaluation set. The numbers in parentheses are the numbers of “Other” class images.

| Manga title | Number of classes | Number of images (Number of “Other” class images) |
|---|---|---|
| ARMS | 14 | 319 (33) |
| Aisazu Niha Irarenai | 11 | 972 (87) |
| Akkera Kanjinchou | 12 | 704 (126) |
| Akuhamu | 9 | 1047 (48) |
| Aosugiru Haru | 12 | 552 (60) |
| Appare Kappore | 11 | 864 (38) |
| Arisa | 11 | 960 (104) |
| BEMADER・P | 16 | 1226 (96) |
| Bakurestu KungFu Girl | 15 | 1036 (121) |
| Belmondo | 13 | 847 (28) |
| Love Hina_vol.14 | 11 | 1168 (43) |

Clustering results are evaluated by V-measure,[14] ARI,[15] and AMI.[16] These values all range from 0.0 to 1.0, and the closer the clustering result is to the correct label, the higher the value. Considering random numbers, calculate the average value of 10 clustering results for each comic book and compare the average values of 11 books.

### 4.1 Evaluation of dimension reduction

First, we investigate the effect of dimension reduction by UMAP on clustering results. In this experiment, VGG16 is used for feature extractor, and DBSCAN is used for clustering method. We compare the clustering results when dimension reduction is not performed and when compression is performed to 256 dimensions, 128 dimensions, 64 dimensions, 32 dimensions, and 2 dimensions using UMAP. The DBSCAN parameter *minPts* is set to 10. The experiment results are shown in Table 2.

Table 2. Comparison of clustering results of character face images with dimension reduction applied.

| | Not reduced (4096d) | 256d | 128d | 64d | 32d | 2d |
|---|---|---|---|---|---|---|
| V-measure | 0.013 | 0.427 | 0.424 | 0.426 | 0.422 | 0.436 |
| ARI | 0.002 | 0.320 | 0.313 | 0.316 | 0.310 | 0.309 |
| AMI | 0.007 | 0.411 | 0.407 | 0.408 | 0.405 | 0.413 |

From these results, it can be seen that applying dimension reduction to the features output from CNN is effective in clustering. It was also shown that the difference in the number of dimensions compressed by UMAP has little effect on the clustering results. Therefore, we compress the dimensions of image features to 2 in the following experiments.

### 4.2 Evaluation of clustering method

Next, we evaluate the effectiveness of the feature extractors and clustering methods. In this experiment, *minPts* is set to 10 for DBSCAN, OPTICS and HDBSCN. The OPTICS parameter "*xi*" that determines the minimum steepness on the reachability plot that constitutes a cluster boundary is set to 0.35. This means an upward point in the reachability plot is defined by the ratio from one point to its successor being at most 1 - *xi*. The parameter "*minSamples*" for HDBSCAN is set to 1. Tables 3, 4 and 5 show the results of applying DBSCAN, OPTICS, and HDBSCAN clustering for the three types of feature extractors.

Table 3. Comparison of character face image clustering results using DBSCAN

| | VGG16 | ResNet50 | ResNet101 |
|---|---|---|---|
| V-measure | 0.436 | 0.637 | 0.634 |
| ARI | 0.309 | 0.505 | 0.483 |
| AMI | 0.413 | 0.619 | 0.616 |

Table 4. Comparison of character face image clustering results using OPTICS

| | VGG16 | ResNet50 | ResNet101 |
|---|---|---|---|
| V-measure | 0.164 | 0.469 | 0.470 |
| ARI | 0.097 | 0.300 | 0.320 |
| AMI | 0.160 | 0.460 | 0.461 |

Table 5. Comparison of character face image clustering results using HDBSCAN

| | VGG16 | ResNet50 | ResNet101 |
|---|---|---|---|
| V-measure | 0.399 | 0.638 | 0.635 |
| ARI | 0.225 | 0.506 | 0.504 |
| AMI | 0.371 | 0.616 | 0.616 |

In comparison of feature extractors, it was confirmed that ResNet50 and ResNet101 showed higher accuracy than VGG16. The reason for this is thought to be that the character description performance improved by the increasing the number of CNN layers. However, the comparison between ResNet101 and ResNet50 did not show a clear improvement. This is considered to be because the character face image features differ between the training set and the evaluation set. In comparison with clustering methods, it was found that when ResNet50 and ResNet101 were used as feature extractors, HDBSCAN showed almost the same accuracy as DBSCAN which adjusted *Eps* values. From this result, it can be said that HDBSCAN can perform useful clustering without complicated parameter settings. On the other hand, in the case of using OPTICS, effective results were not obtained in this experiment.

## 5. CONCLUSION

In this paper, we examined the effectiveness of the feature extractor model and the density-based clustering method for clustering of manga characters. Experimental results show that HDBCSCAN is effective in automating character face classification. In addition, as a general character face image feature extractor, it was confirmed that ResNet50 and

ResNet101 show almost the same performance. As a future subject, learning using only images cannot cope with feature changes in some character face images. To solve this problem, it is necessary to establish a character recognition method that utilizes word information contained in manga simultaneously with images.

## AKNOWLEGEMENT

This work was supported by JSPS KAKENHI Grant Number 17K00511.

## REFERENCES

[1] Nagao, K. and Watanabe, H., “A study on major character identification in comics,” Proc. 2016 The Institute of Electronics, Information and Communication Engineers (IEICE) General Conference, D-21-3 (2016).
[2] Tsubota, K., Ogawa, T., Yamasaki, T., and Aizawa, K., “Character representation adaption in individual manga,” The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 1K2-OS-2b-04 (2018).
[3] Yanagisawa, H., Yamashita, T., and Watanabe, H., “Manga character clustering with DBSCAN using fine-tuned CNN model,” Proc. Volume 11049, International Workshop on Advanced Image Technology (IWAIT) 2019, 110491M (2019).
[4] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L., “ImageNet large scale visual recognition challenge,” International Journal of Computer Vison, 115(3), 211-252 (2015)
[5] Matsui Y., Ito K., Aramaki Y., Fujimoto A., Ogawa T., Yamasaki T. and Aizawa K., “Sketch-based manga retrieval using Manga109 dataset,” Multimedia Tools and Applications, Springer, (2017)
[6] Ogawa T., Otsubo A., Narita R., Matsui Y., Yamasaki T. and Aizawa K., “Object detection for comics using Manga109 annotations,” arXiv:1803.08670, (2018)
[7] Mclnnes L., Healy J. and Melville J, “UMAP: Uniform manifold approximation and projection for dimension reduction,” arXiv:1802.03426, (2018)
[8] Ester M., Kriegel H.P., Sander J. and Xu X., “A density-based algorithm for discovering clusters in large spatical databases with noise,” Proc. 2nd International Conference on Knowledge Discovery and Data Mining, 226-231, (1996)
[9] Soni, N., and Ganatra, N., “AGED (Automatic generation of Eps for DBSCAN),” International Journal of Computer Science and Information Security, 14(5), 536-559, (2016)
[10] Ankerst M., Breunig M.M., Kriegel H.P. and Sander J., “OPTICS: ordering points to identify the clustering structure,” Proc. 1999 ACM SIGMOD International Conference on Management of Data, 49-60, (1999)
[11] Campello R.J.G.B., Moulavi D. and Sander J., “Density-based clustering based on hierarchical density estimates,” Advances in Knowledge Discovery and Data Mining. PAKDD 2013. Lecture Notes in Computer Science, 7819. Springer, Berlin, Heidelberg, (2013)
[12] Simonyan K. and Zisserman A., “Very deep convolutional networks for large-scale image recognition,” arXiv:1409.1556, (2014)
[13] He K., Zhang X., Ren S. and Sun J., “Deep residual learning for image recognition,” arXiv:1512.03385, (2015)
[14] Rosenberg A. and Hirschberg J., “V-measure: a conditional entropy-based external cluster evaluation measure,” Proc. 2007 Joint Conference on Empirical Methods in Natural Language Proceeding and Computational Natural Language Learning (EMNLP-CoNLL), 410-420, (2007)
[15] Hubert, L. and Arabie, P., “Comparing partitions,” Journal of Classification, 2(1), 193-218, (1985)
[16] Vinh N.X., Epps, J. and Bailey J., “Information theoretic measures for clusterings comparison: variants, properties, normalization and correction for chance,” Journal of Machine Learning Research, 11, 2837-2854, (2010)