IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 13, No. 2, June 2024, pp. 2283~2290
ISSN: 2252-8938, DOI: 10.11591/ijai.v13.i2.pp2283-2290

Semi-supervised spectral clustering using shared nearest neighbor for data with different shape and density

Gao YouSheng[1,2], Siti Khatijah Nor Abdul Rahim[1], Raseeda Hamzah[3], Li Ang[1,2], Raihah Aminuddin[3]
[1]College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Shah Alam, Malaysia
[2]College of Information Engineering, Jiujiang Vocational University, Jiu Jiang, China
[3]Computing Sciences Studies, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA (UiTM), Melaka, Malaysia

Article Info

Article history:

Received Aug 10, 2023
Revised Nov 16, 2023
Accepted Jan 6, 2024

Keywords:

K-means
Machine learning
Pairwise constraints
Semi-supervised clustering
Shared nearest neighbor
Spectral clustering

ABSTRACT

In the absence of supervisory information in spectral clustering algorithms, it is difficult to construct suitable similarity graphs for data with complex shapes and varying densities. To address this issue, this paper proposes a semi-supervised spectral clustering algorithm based on shared nearest neighbor (SNN). The proposed algorithm combines the idea of semi-supervised clustering, adding SNN to the calculation of the distance matrix, and using pairwise constraint information to find the relationship between two data points, while providing a portion of supervised information. Comparative experiments were conducted on artificial data sets and University of California Irvine machine learning repository datasets. The experimental results show that the proposed algorithm achieves better clustering results compared to traditional K-means and spectral clustering algorithms.

Corresponding Author:

Raseeda Hamzah
Computing Sciences Studies, College of Computing, Informatics and Mathematics
Universiti Teknologi MARA (UiTM)
Melaka, Malaysia
Email: raseeda@uitm.edu.my

1. INTRODUCTION

Clustering algorithms have been widely used in computer vision, information retrieval, data mining, and other fields for effective data analysis. Clustering analysis belongs to unsupervised learning, which refers to the process of autonomous learning when the information such as data quantity, data attributes, and number of data categories is unknown, and finally constructing partitioning [1]. The traditional K-means clustering algorithm [2] has attracted much attention due to its simplicity and efficiency. K-means algorithm is based on the distribution of convex sample space, but when the sample space is not convex, the algorithm will fall into a local optimal [3]. When the initial value of K-means is selected, an initial cluster is generated. The meaning is basically between the maximum and minimum values of this cluster [4]. After each update of the K-means mean, the cluster can only have a small movement because the mean cannot jump out of the range of the initial cluster [5]. As a result, it is difficult for K-means to obtain the globally optimal cluster and may fall into local optima [6]. To solve this problem, researchers have proposed spectral clustering algorithm [7]. Spectral clustering can make up for this disadvantage of K-means algorithm. The spectral clustering method does not make strong assumptions about the shape of the cluster [8], nor does it have a local optimal solution. Spectral clustering uses the spectral characteristics of the affinity matrix of data samples for dimensionality reduction and uses K-means for clustering in low dimensional spaces [9].

To assist clustering, semi supervised clustering uses less prior information and can be represented by class labels or constraint relationships [10]. Constraint relationships are used to determine whether paired sample data can be classified into one class [11]. In order to apply constraint relationships to clustering, Baumann and Hochbaum [12] proposed paired constraints: must-link (ML) and cannot-link (CL). ML means that two points must be divided into the same class, while CL means that two points must be divided into different classes [13]. In the dataset used for clustering, the degree of difference between various categories is sometimes significant, and the original clustering algorithm is difficult to effectively handle data sets with large differences in density. To address density issue, Wahid and Rao [14] proposed an unsupervised density-based outlier detection algorithm that uses K-nearest neighbour (KNN) and shared nearest neighbor (SNN) to estimate the density, achieving better outlier detection performance. The idea of the SNN algorithm is to first construct a similarity matrix, then process the coefficients of the nearest k neighbors, and then construct a nearest neighbor graph. Only samples with strong connections have links, and then calculate the link strength of all sample points. Sample points with stronger links are more likely to be clustered into one category.

The SNN algorithm only considers the number of the SNN between two data points but does not consider the distance between the two data points. When the spectral clustering algorithm constructs the similarity matrix, it needs to calculate the distance between two points [15]. To improve spectral clustering performance for different densities and shapes, this paper considering the number of SNN and distance between to data points combines SNN information, proposes a semi-supervised spectral clustering algorithm based on SNN. This algorithm can handle data sets with large density differences and uneven densities well, achieving better clustering results.

2. METHOD

The research flow is described in Figure 1 that shows the flowchart of the clustering algorithm and experimental method applied. At the beginning, spectral clustering algorithm, pairwise constraints, and SNN similarity are introduced. Then, the problems of spectral clustering are analyzed. To overcome the shortcomings of spectral clustering, semi-supervised spectral clustering algorithm using shared nearest neighbor (SSCSNN) is proposed. Then select accuracy as the evaluation indicator for the clustering algorithm. Finally, the performance of the clustering algorithm was experimentally verified using artificial datasets and University of California Irvine (UCI) datasets.

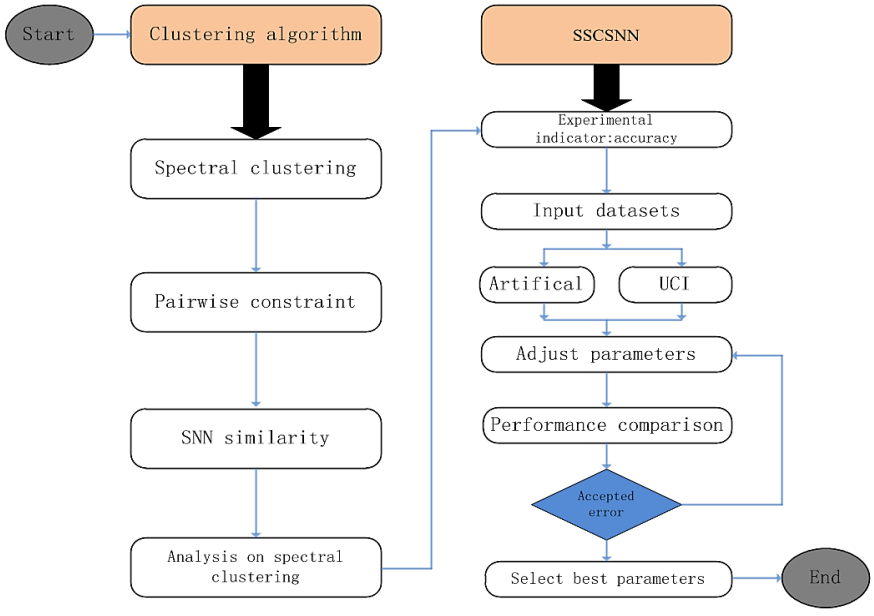

Figure 1. Research flow

2.1. Spectral clustering

Spectral clustering is a clustering method based on graph theory, which treats each sample in the dataset to be clustered as a vertex in the graph [16]. These vertices are connected. The size of the connected

weights indicates the similarity between these samples. Vertices of the same class have a high degree of similarity [17]. In graph theory, the weight of the edges connecting them is large among vertices of the same class, while the weight of the edges connecting them between vertices that are not in the same class is small [18]. The goal of spectral clustering is to find a method of cutting graphs so that the weights within each subgraph after cutting are large, while the weights between subgraphs are small [19]. Figure 2 shows how to cut a graph into two subgraphs.

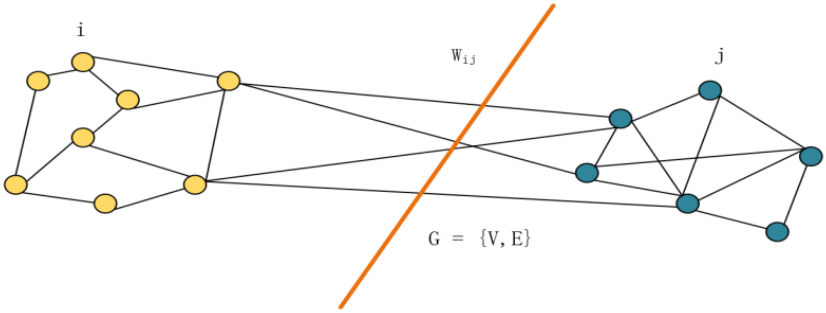

Figure 2. Graphs cutting

Given a data set, the edge weights W between the vertices are assigned based on the similarity between the samples, resulting in an undirected weighted graph $G=(V,E)$ based on the sample similarity. Where $V=\{v_1,v_2,...,v_n\}$ represents a vertex and E represents a set of edges. The weight of the edges connecting two vertices v_i and v_j is recorded as w_{ij}, and their similarity is represented by s_{ij}. The larger the value of this similarity is, the more similar they are. W is a symmetric matrix, that is, $W=W^T$. For a vertex $d_i,i=1,2,...n$, define degrees as in (1),

$$d_i = \sum_{j=1}^{n} w_{ij} \quad (1)$$

The degree matrix is defined as a diagonal matrix composed of n degrees. The similarity measurement of each vertex in the graph is mainly based on distance, which means that the closer the distance between two points in space is, the more similar they are, and the farther the distance is, the less similar they are [20]. That is, similarity is inversely proportional to distance. Therefore, as long as the metric space used has this property, it can be used as a measure of similarity. Euclidean distance is a widely used similarity measure method [21]. x_i and x_j represent a data point vector of the d dimension. In general, the closer the distance between data points is, the higher the similarity they are. Through the similarity matrix W and the degree matrix D, the Laplace matrix L can be calculated as in (2),

$$L = D - W \quad (2)$$

After obtaining the Laplace matrix, we can perform normalization operations to calculate the normalized graph Laplace matrix as in (3),

$$L_{sym} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} \quad (3)$$

2.2. Pairwise constraints

General clustering algorithms are unsupervised, that is users do not interfere in the clustering process, and the clustering results are only a natural segmentation of data objects and cannot represent user preferences [22]. However, actual clustering analysis is intended to serve specific users, and clustering algorithms should be able to consider user preferences [23]. The tendency of people in clustering can often be expressed as follows: certain two data objects should be classified into one category or certain two objects should not be classified into one category. Baumann *et al.* [24] introduced the concept of paired constraints. Paired constraints are divided into ML and CL. Two sample objects labeled with ML represent two samples that must be assigned to the same cluster, while two sample objects labeled with CL represent two samples that must be assigned to different clusters. Figure 3 shows a negative association constraint relationship between x_i and x_j, where x_i represents the blue circle sample data points in cluster 1, x_j represents the yellow triangle sample data points in cluster 2, and a solid arrow indicates that x_i and x_j belong to a negative association constraint relationship.

2.3. Shared nearest neighbor

SNN is based on the fact that if two points are similar to some of the same points, they are similar even if a direct similarity measure cannot indicate them [25]. More specifically, as long as two objects are in each other's nearest neighbor's table, SNN similarity is the number of nearest neighbors they share [26]. Suppose that two objects A and B have 8 closest neighbors, and the two objects contain each other. Among these closest neighbors, 4 are shared by A and B, so the SNN similarity between the two objects is 4. The SNN similarity graph between objects is called the SNN similarity graph. Because the SNN similarity between many objects is 0, the SNN similarity graph is very sparse. Figure 4 shows illustrates the SNN similarity between 2 points.

SNN similarity solves some problems when using direct similarity. First, it can handle situations where one object happens to be relatively close to the other object, but the two belong to different clusters. In this case, two objects generally do not contain many shared neighbors, so their SNN density is very low. In another case, when processing clusters with different densities, the SNN density will be appropriately scaled based on the density of the cluster, as the similarity between a pair of objects does not depend on the distance between the two, but rather on their shared neighbors.

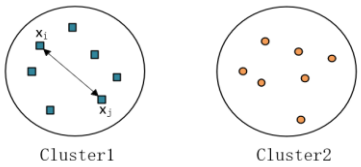

Figure 3. Positive association constraint of paired sample points

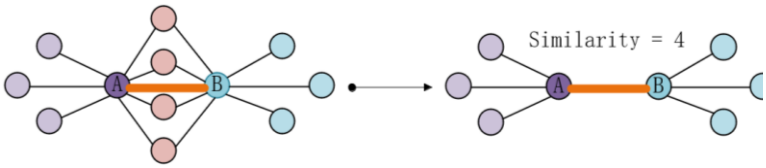

Figure 4. Calculation of SNN similarity between two points

2.4. Semi-supervised spectral clustering using shared nearest neighbor

To cluster data sets with different densities and shapes, this paper proposes a SSCSNN. In order to combine SNN information with spectral clustering, this article adds the idea of SNN, which is a relatively new similarity measurement method that can adaptively adjust the similarity between data points based on their local density. Researchers have found that if two points belong to the same cluster, they should be located in the same region, with a relatively high density, and there will be many overlapping points between the two points. To represent the glue state between these two points, it is represented by a SNN. *SNN* (x_i,x_j) represents the number of neighbors shared by sample points x_i and x_j in the k-neighborhood. *SNN* (x_i,x_j) reflects the local density of x_i and x_j points, which is helpful for the correct division of sample points. SNN is a locally adaptive density data that can be used to improve the similarity between data points. The similarity matrix calculation method in this article is based on the combination of SNN algorithm and Euclidean distance. The formula is as (4),

$$Dist(x_i, x_j) = (1 - \frac{1}{1+e^{-SNN(x_i,y_j)}})\sqrt{\sum_{l=1}^{d}(x_{il} - x_{jl})^2} \quad (4)$$

Input: Dataset $X=\{x_1,x_2,...,x_n\}$, cluster number k, constraint matrix P
Output: Clustering results
Step 1: Calculate the distance between samples according to (4), and then modify the distance matrix according to the constraint matrix P.

$$Dist(x_i, y_j) = \begin{cases} 0, if(x_i, y_j) \in Must - Link \\ \infty, if(x_i, y_j) \in Cannot - Link \end{cases} \quad (5)$$

Step 2: Establish the similarity matrix W of the sample set based on (5), and define the cluster number k in advance.
Step 3: Calculate the degree matrix D
Step 4: Calculate and standardize the Laplace matrix L
Step 5: Calculate the eigenvector f corresponding to the smallest k eigenvalues of the normalized Laplace matrix
Step 6: Standardize the matrix composed of each feature vector f by row to form a feature matrix
Step 7: Uses K-means to cluster each row in F as a k-dimensional data point

3. RESULTS AND DISCUSSION

The experiment was conducted on four artificial datasets and UCI machine learning repository datasets. The experimental environment was 11th Gen Intel (R) Core (TM) i5-11400H @ 2.70 GHz, the memory was 24 GB DDR4 3200 Hz, and the programming environment was Python 3.8.0. The test was conducted on the Windows 10 operating system.

3.1. Clustering indicators

To compare and analyze clustering results, the accuracy clustering index was used in the experiment to measure. Accuracy is widely used to evaluate the performance of clustering algorithms. The accuracy calculation is as (6),

$$ACC = \frac{1}{n}\sum_{i=1}^{c}|T_i \cap P_i| \quad (6)$$

$T_i=\{T_1,T_2,...,T_c\}$ indicates that the original n data contains true c categories, and $P_i=\{P_1,P_2,...,P_c\}$ indicates c prediction categories of the n data after clustering. T_i represents the number of points included in the ith category. P_i represents the data points contained in the ith category after clustering, and $|P_i|$ represents the number of points contained in the set P. The clustering index value is 0-1, and the larger the value is, the better the clustering effect is.

3.2. Dataset

To visualize the clustering algorithm proposed in this article in a two-dimensional plane, four two-dimensional artificial datasets are used to test the clustering effects of K-means, spectral clustering, and SSCSNN algorithms respectively. Figure 5 shows 4 datasets with different shape. Dataset1 contains 7 point sets with different densities and shapes composed of 788 points. Dataset2, dataset3, and dataset4 are directly generated by the Python based open source machine learning toolkit scikit-learn. Dataset2 is generated by the make_circles function in the scikit-learn toolkit, containing 1,600 points and two classes. Dataset3 is generated by the make_moons function in the scikit-learn toolkit, containing 1,000 points and two classes. Dataset4 is generated by the make_blobs function in the scikit-learn toolkit, containing 1,600 points and 5 classes. Like the UCI datasets, these four artificial datasets all have different densities and shapes. To verify the effectiveness of the algorithm, 4 real-world datasets were also selected from UCI benchmark data. The parameters of the UCI datasets and artificial datasets are shown in Table 1.

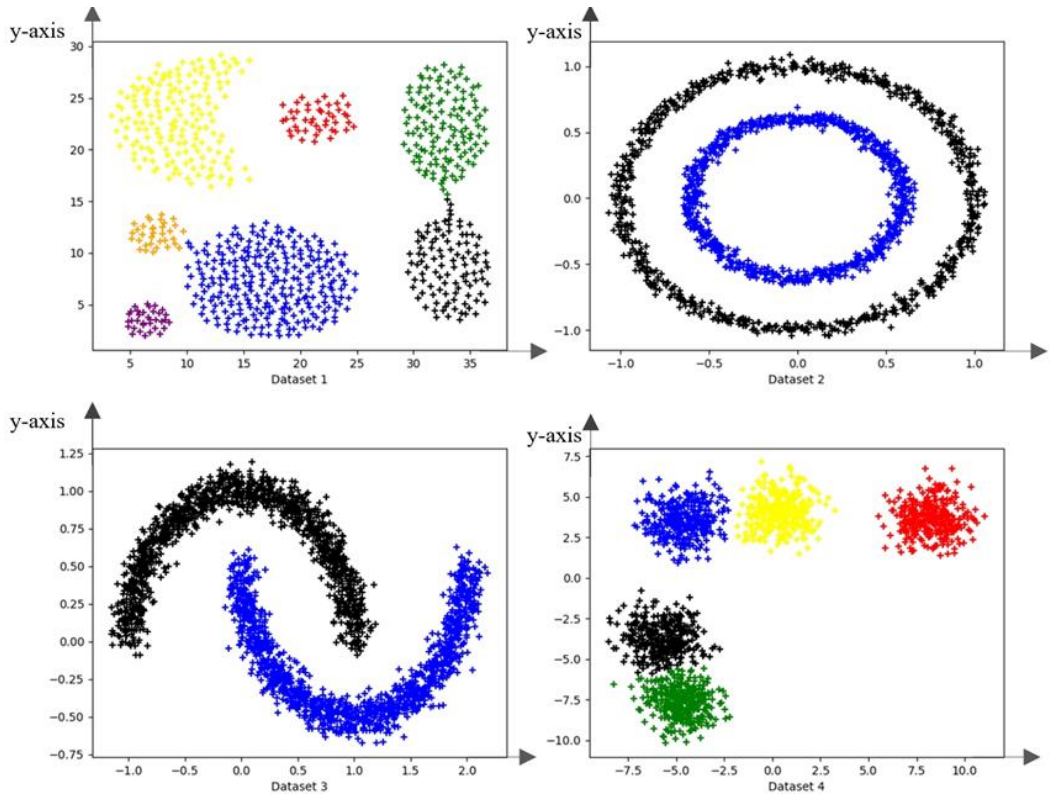

Figure 5. Datasets with different shape

Table 1. Parameters of UCI datasets and artificial datasets

Datasets	Instances	Number of features	Number of classes
Dataset1	788	2	7
Dataset2	1,600	2	2
Dataset3	1,000	2	2
Dataset4	3,200	2	5
Glass	214	9	6
Iris	150	4	3
Seeds	210	7	3
Wine	178	13	3

3.3. Experimental results and analysis

The clustering experimental results of the artificial datasets are shown in Table 2. The artificial dataset experimental results are calculated based on 10 repeated independent experiments. The indicator for evaluating clustering performance is clustering accuracy. It is necessary to calculate the proportion of the classes of each instance in the dataset that match the true labels.

In Dataset1, the average accuracy of K-means is 91.37%, and the average accuracy of SSCSNN is slightly 0.12% higher than the average accuracy of spectral clustering. In dataset2 and dataset3, the average accuracy of K-means is not as good as spectral clustering and SSCSNN with an average accuracy of 100%. In dataset4, the average accuracy of K-measure is 69.34%, and the average accuracy of SSCSNN is 3.73% higher than the average accuracy of spectral clustering. From the overall experimental results, SSCSNN clustering is superior to K-means and spectral clustering. The clustering experimental results of the artificial dataset are shown in Table 3.

The experimental results of UCI dataset are calculated based on 10 repeated independent experiments, and the indicator for evaluating clustering performance is clustering accuracy. In the glass dataset, K-means and spectral clustering algorithms have average performance, while SSCSNN clustering is superior to these two algorithms. In the iris dataset, the accuracy rate of K-means clustering algorithm is stable at 50%, and the accuracy rates of spectral clustering and SSCSNN clustering algorithms have reached 100%. In the seeds dataset, the accuracy rate of K-means clustering algorithm is stable at 50%, and the accuracy rate of SSCSNN clustering algorithm is 5.71% higher than spectral clustering algorithm, with an accuracy rate of 95.23%. In the wine dataset, the three algorithms perform relatively generally, but the SSCSNN clustering algorithm is still significantly superior to the other two clustering algorithms. From the experimental results of UCI data sets, the SSCSNN clustering algorithm is generally better than K-means and spectral clustering algorithms.

Table 2. Clustering results of artificial datasets

	Accuracy (%)		
	K-means	Spectral	SSCSNN
Dataset1	88.95	99.62	99.74
Dataset2	50.25	100	100
Dataset3	64.32	100	100
Dataset4	69.34	94.64	98.37

Table 3. Clustering results of UCI datasets

	Accuracy (%)		
	K-means	Spectral	SSCSNN
Glass	61.12	72.42	80.84
Iris	50.00	100	100
Seeds	50.00	89.52	95.23
Wine	52.60	64.60	73.03

4. CONCLUSION

This paper proposes a semi-supervised spectral clustering algorithm based on SNN. The distance matrix in SSCSNN algorithm incorporates the SNN information and paired constraint information from the SNN algorithm. Experimental results show that this algorithm performs better than K-means and spectral clustering algorithms on both synthetic and UCI datasets.

ACKNOWLEDGEMENTS

The researchers would like to express the College of Computing, Informatics and Mathematics, Universiti Teknologi MARA (UiTM) for their support.

REFERENCES

[1] J. Xie, M. Wang, X. Lu, X. Liu, and P. W. Grant, "DP-k-modes: A self-tuning k-modes clustering algorithm," *Pattern Recognition Letters*, vol. 158, pp. 117–124, 2022, doi: 10.1016/j.patrec.2022.04.026.

[2] B. Chong, "K-means clustering algorithm: a brief review," *Academic Journal of Computing & Information Science*, vol. 4, no. 5, pp. 37–40, 2021, doi: 10.25236/AJCIS.2021.040506.

[3] H. Hu, J. Liu, X. Zhang, and M. Fang, "An effective and adaptable k-means algorithm for big data cluster analysis," *Pattern Recognition*, vol. 139, 2023, doi: 10.1016/j.patcog.2023.109404.

[4] R. Ghezelbash, A. Maghsoudi, M. Shamekhi, B. Pradhan, and M. Daviran, "Genetic algorithm to optimize the SVM and K-means algorithms for mapping of mineral prospectivity," *Neural Computing and Applications*, vol. 35, no. 1, pp. 719–733, 2023, doi: 10.1007/s00521-022-07766-5.

[5] J. Chen, S. Xie, H. Jiang, H. Yang, and F. Nie, "A novel k-means framework via constrained relaxation and spectral rotation," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2023, doi: 10.1109/TNNLS.2023.3282938.

[6] K. G. Dhal, R. Rai, A. Das, and T. K. Ghosh, "Hybridization of sine-cosine algorithm with k-means for pathology image clustering," *Communications in Computer and Information Science*, vol. 1695 CCIS, pp. 76–86, 2022, doi: 10.1007/978-3-031-22485-0_8.

[7] F. S. Passino, N. A. Heard, and P. Rubin-Delanchy, "Spectral Clustering on Spherical Coordinates Under the Degree-Corrected Stochastic Blockmodel," *Technometrics*, vol. 64, no. 3, pp. 346–357, 2022, doi: 10.1080/00401706.2021.2008503.

[8] Z. Zhou and A. A. Amini, "Analysis of spectral clustering algorithms for community detection: The general bipartite setting," *Journal of Machine Learning Research*, vol. 20, pp. 1–47, 2019.

[9] A. A. Qaffas, M. A. B. Hajkacem, C. E. B. Ncir, and O. Nasraoui, "Interpretable Multi-Criteria ABC Analysis Based on Semi-Supervised Clustering and Explainable Artificial Intelligence," *IEEE Access*, vol. 11, pp. 43778–43792, 2023, doi: 10.1109/ACCESS.2023.3272403.

[10] L. Peng *et al.*, "GRLC: Graph Representation Learning With Constraints," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2023, doi: 10.1109/TNNLS.2022.3230979.

[11] H. Zhu, S. Klus, and T. Sahai, "A Dynamic Mode Decomposition Approach for Decentralized Spectral Clustering of Graphs," in *2022 IEEE Conference on Control Technology and Applications (CCTA)*, 2022, pp. 1202–1207, doi: 10.1109/CCTA49430.2022.9966092.

[12] P. Baumann and D. Hochbaum, "A k-Means Algorithm for Clustering with Soft Must-link and Cannot-link Constraints," in *Proceedings of the 11th International Conference on Pattern Recognition Applications and Methods*, 2022, pp. 195–202, doi: 10.5220/0010800000003122.

[13] S. Anvari, M. A. Azgomi, M. R. E. Dishabi, and M. Maheri, "Weighted K-nearest neighbors classification based on Whale optimization algorithm," *Iranian Journal of Fuzzy Systems*, vol. 20, no. 3, pp. 61–74, 2023, doi: 10.22111/ijfs.2023.7639.

[14] A. Wahid and A. C. S. Rao, "An Outlier Detection Algorithm based on KNN-kernel Density Estimation," in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020, pp. 1–8, doi: 10.1109/IJCNN48605.2020.9207033.

[15] M. Kleindessner, S. Samadi, P. Awasthi, and J. Morgenstern, "Guarantees for spectral clustering with fairness constraints," in *36th International Conference on Machine Learning, ICML 2019*, 2019, pp. 6004–6030.

[16] R. Janani and S. Vijayarani, "Text document clustering using Spectral Clustering algorithm with Particle Swarm Optimization," *Expert Systems with Applications*, vol. 134, pp. 192–200, 2019, doi: 10.1016/j.eswa.2019.05.030.

[17] K. Berahmand, S. Haghani, M. Rostami, and Y. Li, "A new attributed graph clustering by using label propagation in complex networks," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 5, pp. 1869–1883, 2022, doi: 10.1016/j.jksuci.2020.08.013.

[18] T. Richthammer, "Bunkbed conjecture for complete bipartite graphs and related classes of graphs," *ArXiv-Mathematics*, pp. 1–10, 2022.

[19] B. Chander and K. Gopalakrishnan, "Data clustering using unsupervised machine learning," in *Statistical Modeling in Machine Learning*, Cambridge: Academic Press, 2023, pp. 179–204, doi: 10.1016/B978-0-323-91776-6.00015-4.

[20] K. K. Sharma and A. Seal, "Multi-view spectral clustering for uncertain objects," *Information Sciences*, vol. 547, pp. 723–745, 2021, doi: 10.1016/j.ins.2020.08.080.

[21] K. K. Sharma, A. Seal, E. Herrera-Viedma, and O. Krejcar, "An enhanced spectral clustering algorithm with s-distance," *Symmetry*, vol. 13, no. 4, pp. 1–17, 2021, doi: 10.3390/sym13040596.

[22] W. Yu, L. Xing, F. Nie, and X. Li, "An efficient semi-supervised balanced cut with hard pairwise constraints and partial labels," *Knowledge-Based Systems*, vol. 276, 2023, doi: 10.1016/j.knosys.2023.110747.

[23] A. El Moussawi, A. Giacometti, N. Labroche, and A. Soulet, "MAPK-means: A clustering algorithm with quantitative preferences on attributes," *Intelligent Data Analysis*, vol. 24, no. 2, pp. 459–489, 2020, doi: 10.3233/IDA-184468.

[24] P. Baumann, "A binary linear programming-based k-means algorithm for clustering with must-link and cannot-link constraints," in *2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 2020, pp. 324–328, doi: 10.1109/IEEM45057.2020.9309775.

[25] I. D. Wahyono *et al.*, "Shared nearest neighbour in text mining for classification material in online learning using mobile application," *International Journal of Interactive Mobile Technologies*, vol. 16, no. 4, pp. 159–168, 2022, doi: 10.3991/ijim.v16i04.28991.

[26] S. M. Islam, S. M. M. Hossain, and S. Ray, "DTI-SNNFRA: Drug-target interaction prediction by shared nearest neighbors and fuzzy-rough approximation," *PLOS ONE*, vol. 16, no. 2, pp. 1–19, 2021, doi: 10.1371/journal.pone.0246920.

BIOGRAPHIES OF AUTHORS

Gao YouSheng has been a lecturer in Jiujiang Vocational University since September 2020. He holds a master degree of engineering (M.Eng.) in Computer Engineering. He has been a Ph.D. candidate in Universiti Teknologi MARA since October 2022. He is currently a researcher in the direction of artificial intelligence in the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor. He can be contacted at email: 2022287318@student.uitm.edu.my.

Siti Khatijah Nor Abdul Rahim is a senior lecturer at the College of Computing, Informatics and Media, Universiti Teknologi MARA, Shah Alam, Selangor. She obtained her Ph.D. in Computer Science from the University of Nottingham in 2015. Before that, she received her bachelor's degree in Computer Science and M.Sc. of Science (Computer Science) from Universiti Sains Malaysia (USM), Penang, Malaysia, in 2001 and 2003, respectively. Her main research areas lie in the area of computational intelligence, granular computing, scheduling, optimization, and applied computing. She can be contacted at email: sitik781@uitm.edu.my.

Raseeda Hamzah has been a lecturer in Universiti Teknologi MARA since 12 August 2016. She is currently a senior lecturer at the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor. Secured her Ph.D. in Information Technology and Quantitative Sciences, Universiti Teknologi Mara, Cawangan Shah Alam. Her area of expertise is machine learning, artificial intelligence, and digital signal processing. She is also actively doing research in Urban Farming and IoT. She can be contacted at email: raseeda@uitm.edu.my

Li Ang has been a lecturer in Jiujiang Vocational University since September 2018. He is currently a Ph.D. candidate at the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor. He received Master of Engineering (M.Eng.) in Information and Communication Engineering, Guangdong University of Technology. His research areas of interest include computer vision, artificial intelligent, and digital signal processing. He can be contacted at email: 2022667284@student.uitm.edu.my.

Raihah Aminuddin is a senior lecturer at the College of Computing, Informatics, and Mathematics, Universiti TFieknologi MARA, Jasin, Melaka, Malaysia. She obtained her Ph.D. in Computer Science from the University of Sheffield, United Kingdom, in 2019. Before that, she received her bachelor's degree in Intelligent Systems and Master of Science in Information Technology from Universiti Teknologi MARA, Shah Alam, Malaysia, in 2011 and 2013, respectively. Her areas of expertise are machine learning, artificial intelligence, image processing, and big data. She can be contacted at email: raihah1@uitm.edu.my.