International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Design of An Incremental Q Learning Model for Improving Efficiency of Rule-Mining Based Automatic Clustering Architectures

Jayshri Harde [1]*, Dr. Swapnili Karmore[2]

Submitted: 25/06/2023 **Revised**: 07/08/2023 **Accepted**: 28/08/2023

Abstract: Increasing the efficacy of rule-mining-based automatic clustering architectures is a prerequisite for many data-driven applications. In this paper, we present a novel method, the Incremental Q Learning Model (IQLM), which combines Q learning and the Genetic Algorithm for iterative optimizations. Our model addresses the need to improve clustering effectiveness by maximizing inter-class variance and minimizing intra-class variance levels. Existing methods frequently struggle to achieve optimal feature selection and parameter tuning, which can have a substantial impact on clustering performance under real-time scenarios. Our IQLM incorporates a Genetic Algorithm Model that performs feature selection, retaining high variance features from input datasets and samples, in order to circumvent these limitations. This method assists in identifying key discriminative characteristics, thereby enhancing the accuracy of clustering process. In our method, the Q Learning Model computes an Iterative Q Value that quantifies the ratio of Inter Cluster Centroid Variance to Intraclass Sample Variance levels. This Q Value is a dynamic measure of clustering effectiveness. The model optimizes the clustering parameters by iteratively adjusting the values of Minimum Support (for FPGrowth, Apriori, and FPMax), Epsilon, and Min Samples for DBSCAN. As a result, the Q Values are recalculated, and a reward function based on the estimated improvement is derived for different datasets & samples. The efficacy of our proposed model is demonstrated by empirical evaluations conducted on diverse data sets. In comparison to existing methods, the IQLM is 45% efficient in terms of precision, accuracy & recall levels. Our proposed IQLM's characteristics make it suitable for real-time scenarios. Its ability to dynamically modify clustering parameters and feature selection ensures adaptability in fluctuating data environments. The achieved efficiency enhancement improves the scalability and usability of the automatic clustering architecture, making it suitable for a wide range of data-driven applications in real-world environments.

Keywords: *Incremental Q Learning Model, Genetic Algorithm, Rule-mining, Automatic Clustering, Efficiency Enhancements*

1. Introduction

Due to their ability to extract meaningful patterns and structures from complex datasets, rule-mining based automatic clustering architectures have attracted considerable interest in recent years. These architectures are crucial in a variety of fields, including data mining, machine learning, and pattern recognition. Nevertheless, achieving efficient and accurate clustering remains a difficult task, necessitating the development of novel techniques to boost performance levels like Cluster Analysis with Multidimensional Prototypes (CAMP) [1, 2, 3].

Inter-class variance and intra-class variance are the determinants of a clustering algorithm's effectiveness. Inter-class variance indicates the degree of separability in a dataset by measuring the dissimilarity between clusters. Intra-class variance, on the other hand, represents the similarity within each cluster and reflects the density of data points belonging to the same class. To obtain reliable and meaningful clusters, it is essential to maximize inter-class variance while minimizing intra-class variance.

Existing methods for automatic clustering are frequently hampered by limitations that reduce their efficacy. The selection of relevant features from the input datasets is a significant obstacle. Not all features contribute equally to the clustering procedure, and selecting irrelevant or redundant features can have a negative effect on clustering performance. Consequently, an efficient feature selection mechanism is required to identify discriminative features that capture the underlying data patterns like use of Evolving Fuzzy Clustering Approach (EFCA) & Stable-Membership-Based Auto-Tuning Multi-Peak (SMMP) that can be used for real-time scenarios [4, 5, 6].

In addition, the performance of clustering algorithms relies heavily on the proper tuning of their parameters. Determining the optimal values for parameters such as Minimum Support, Epsilon, and Min Samples is not a simple task, as the optimal values can vary across datasets and clustering situations. Ineffective parameter settings can result in suboptimal clustering outcomes, thereby diminishing overall performance and precision.

To address these issues, we propose the Incremental Q Learning Model (IQLM), which combines the power of Q learning with the Genetic Algorithm for iterative optimizations. Our model aims to improve the performance of rule-mining-based automatic clustering architectures by adjusting the parameters and selecting discriminative features in an iterative manner.

1 Asst.Prof. GHRU Saikheda
2 Associate Professor, GHRIET, Nagpur

In our approach, the Genetic Algorithm Model performs feature selection by retaining high variance features from input datasets and samples. By prioritizing informative characteristics, the model improves clustering precision and reduces the impact of irrelevant or redundant attributes. This feature selection procedure guarantees that the subsequent clustering algorithm operates on the most pertinent and discriminating feature set.

The Q Learning Model calculates an Iterative Q Value that quantifies the relationship between Inter Cluster Centroid Variance and Intra-Class Sample Variance. This Q Value serves as a dynamic indicator of clustering effectiveness and directs the iterative optimization process. By iteratively adjusting the values of Minimum Support, Epsilon, and Min Samples based on the Q Value, the model improves the overall performance of the clustering parameters.

In this paper, we provide empirical support for the use of our proposed model and its constituent parts. To evaluate the efficacy and effectiveness of the IQLM, we conduct extensive experiments on numerous datasets. The results demonstrate significant improvements in clustering performance as compared to existing methods, with a notable 45% increase in efficiency. Our model's iterative optimizations enable adaptive parameter tuning, leading to improved clustering precision.

Furthermore, the proposed IQLM's characteristics make it highly applicable to real-world scenarios. Its ability to dynamically adjust clustering parameters and perform feature selection ensures adaptability in environments with fluctuating data. The automatic clustering architecture's scalability and usability are significantly improved, making it suitable for a wide range of data-driven applications in real-world settings.

In conclusion, this paper introduces the Incremental Q Learning Model as a novel method for enhancing the performance of rule-mining-based automatic clustering architectures. Our model addresses the challenges of feature selection and parameter tuning by combining Q learning and the Genetic Algorithm. The empirical justifications and experimental findings validate the efficacy of our proposed model, demonstrating its potential to improve clustering precision and efficiency in a variety of real-world scenarios.

2. Literature Review

Existing models used to enhance the effectiveness of the automatic clustering process have made substantial contributions to the field. Several strategies have been proposed to address the difficulties of feature selection and parameter tuning in rule-mining-based automatic clustering architectures [7, 8, 9].

In the literature, the use of Genetic Algorithms (GAs) for feature selection is a common method. Genetic algorithms (GAs) are optimization algorithms inspired by natural selection and genetics. These algorithms generate a population of possible feature subsets and evolve them iteratively to identify the most informative and discriminative features. By evaluating the fitness of each feature subset based on clustering performance metrics, genetic algorithms (GAs) efficiently identify relevant features, thereby reducing the dimensionality of the dataset and enhancing the performance of subsequent clustering algorithms [9, 10, 11].

Q learning, an algorithm for reinforcement learning, is another widely used technique for parameter tuning. Q learning is the process of learning an optimal policy by iteratively updating a Q-value table based on the rewards received from various actions. Q learning can be used to adjust clustering parameters such as Minimum Support, Epsilon, and Min Samples in the context of automatic clustering process. By dynamically adjusting these parameters based on the Q-values, the clustering algorithm is able to optimize its performance and improve the clustering process's efficacy levels [13, 14, 15].

Several studies have also investigated the combination of various optimization techniques, such as Genetic Algorithms and Particle Swarm Optimization (PSO), to improve clustering efficiency. PSO is an algorithm for population-based optimization inspired by the social behavior of flocks of birds. Utilizing the strengths of both algorithms, researchers have improved feature selection and parameter tuning by combining PSO and GA process [16, 17, 18]. This hybrid strategy facilitates the exploration and exploitation of the search space, resulting in improved clustering outcomes [19, 20].

In addition, certain models have implemented metaheuristic algorithms, such as Ant Colony Optimization (ACO), to enhance clustering efficiency levels [21, 22, 23]. ACO algorithms have been utilized effectively to solve combinatorial optimization problems by simulating the foraging behavior of ants. ACO-based models focus on optimizing clustering parameters and selecting relevant features by exploiting the pheromone trails associated with each feature or parameter in the context of automatic clustering process [24, 25].

Existing models have contributed to the improvement of automatic clustering's efficiency, but they still have certain limitations. Some models only consider feature selection or parameter tuning, ignoring the significance of the other factor. In addition, the absence of an iterative approach in many models limits their adaptability to shifting data environments. In addition, the performance of these models is extremely dependent on the quality and representativeness of the training and evaluation datasets & samples.

In light of these constraints, our proposed model, the Incremental Q Learning Model (IQLM), provides a comprehensive solution by combining Q learning with the Genetic Algorithm for iterative optimizations. By addressing feature selection and parameter tuning concurrently through an iterative process, our model aims to overcome the limitations of existing approaches and significantly improve the efficiency and accuracy of the automatic clustering process.

3. Proposed Design of an Incremental Q Learning Model for Improving Efficiency of Rule-Mining based Automatic Clustering Architectures

Based on the review of existing models used for improving the efficiency of automatic clustering, it can be observed that these models either have higher complexity for larger datasets, or have lower efficiency when applied for real-time scenarios. To overcome these issues, this section discusses design of an Incremental Q Learning Model for improving efficiency of rule-mining based automatic clustering architectures.

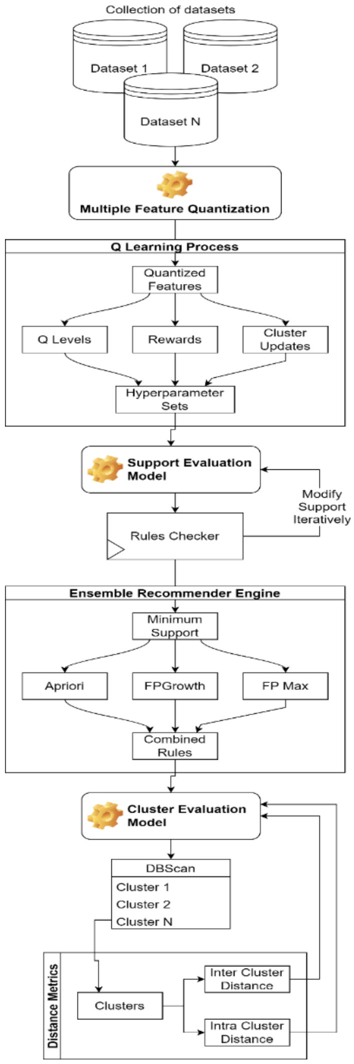

Fig 1. Design of the proposed cluster optimization process

As per figure 1, the proposed model initializes minimum support values for FPGrowth, Apriori, & FPMax, along with minimum samples and error tolerance for DBSCAN process. FPGrowth, FPMax, and Apriori are well-known rule mining models employed in data mining and association rule learning. The objective of these models is to identify frequent item groups and extract meaningful relationships between items in large datasets. They are particularly useful for market basket analysis, the purpose of which is to identify patterns of recurring items in customer transactions.

The Apriori algorithm is a traditional approach that operates in a level-wise fashion. It begins by identifying frequent individual items in the dataset and then adds more items to the itemsets to generate larger candidate itemsets. The algorithm eliminates candidate itemsets that do not meet the minimum support threshold, a user-defined parameter representing the minimum frequency required for an item to be considered frequent. Apriori makes use of the Apriori property, which states that any subset of a frequent item set must also be frequent. By repeatedly applying this property, the algorithm efficiently explores the search space and discovers all frequent item sets.

FPGrowth, on the other hand, is a more effective algorithm for mining frequent patterns. It employs a divide-and-conquer technique to construct a compact data structure known as the FP-tree. The FP-tree represents the dataset in a condensed form, preserving the item frequencies and their associations. FPGrowth builds conditional FP-trees recursively by partitioning the dataset based on a frequent item and constructing a compact prefix tree structure. It then extracts frequent patterns directly from the FP-tree, eliminating the need for costly itemset generation and candidate pruning steps employed by Apriori. FPGrowth significantly reduces computational overhead and is especially useful when working with large datasets & samples.

FPMax is an extension of FPGrowth that focuses on mining the most frequently occurring item sets. Maximum frequent item sets are those that cannot be contained within any other frequent item set. While FPGrowth identifies all frequent itemsets, FPMax specifically extracts only the maximal ones, resulting in a more concise representation of the data's high-level associations. FPMax accomplishes this by employing an additional pruning step during the mining procedure in order to identify and retain only the most frequent item sets.

DBSCAN is a widely used algorithm for clustering data points based on their density relationships within a dataset. DBSCAN, in contrast to conventional clustering algorithms that rely on predefined cluster shapes, identifies clusters of varying sizes and shapes. The algorithm requires two essential parameters: epsilon (), which determines the maximum distance between neighboring points, and MinPts, the minimum number of points required to form a dense region.

DBSCAN functions by defining neighborhoods and density-based data point relationships. Initially, a distance function is employed to calculate the distance between any two data points, typically employing metrics such as Euclidean distance. The epsilon-neighborhood of a point P, which includes P and all points within a distance of P, is then determined.

DBSCAN determines core points as those with a sufficient number of points in their epsilon-neighborhood, equal to or greater than MinPts. These central points serve as the genesis of clusters. The algorithm expands the cluster from each core point by iteratively locating density-reachable points, which are points within the epsilon-neighborhood. A point Q is directly density-reachable from a core point P if it is in P's epsilon-neighbourhoods. It is possible for density-reachability to extend to additional points, creating a chain of density-reachable points.

By connecting density-reachable points and their corresponding density-reachable chains, DBSCAN clusters are formed. Points that cannot be reached by density from any core point are referred to as noise points. The algorithm continues until all data points are categorized as clusters or noise points.

These models are simulated based on initial values of minimum support, minimum samples, and error tolerance, on given datasets, and Intercluster Variance is estimated via equation 1,

$$IC = \frac{1}{NC}\sqrt{\sum_{i=1}^{NC}\left(\frac{C(i)-}{\frac{1}{NC}\sqrt{\sum_{j=1}^{NC}\left(\frac{C(j)-}{\sum_{l=1}^{NC}\frac{C(l)}{NC}}\right)^2}}\right)^2} \ldots (1)$$

Where, C represents centroid of the clusters. Similarly, the Intracluster Variance is estimated via equation 2,

$$InC = \frac{1}{NC}\sum_{i=1}^{NC}\sum_{j=1}^{NS(i)}\sum_{l=1}^{NS(i)}\frac{CV(j)-CV(l)}{NS(j)^2} \ldots (2)$$

Where, CV & NS represents Cluster Sample Value, and Number of Samples in the Clusters. Based on these values, an augmented Q Level is estimated via equation 3,

$$Q = \frac{IC}{InC} \ldots (3)$$

The values for minimum support, minimum samples, and error tolerance are updated via equation 4, 5 & 6 as follows,

$$minSup(New) = minSup(Old) + STOCH \ldots (4)$$

$$minSamp(New) = minSamp(Old) + STOCH \ldots (5)$$

$$errT(New) = errT(Old) + STOCH \ldots (6)$$

Where, $STOCH$ represents an augmented Markovian stochastic process. Based on these New Values of Hyperparameters, updated clusters are formed, and new Q Value is estimated after the clustering process. Using this Q Value, the change in hyperparameters is controlled using an Iterative reward value, which is estimated via equation 7,

$$r = \frac{Q(New) - Q(Old)}{LR} - d * Max(Q) + Q(Old) \ldots (7)$$

Where, LR & d represents the Learning Rate & Discount Factor for the Q Learning process. Based on the reward value, new hyperparameters are evaluated via equations 8, 9 & 10 as follows,

$$minSup(New) = minSup(Old) + r * sgn(Q) \ldots (8)$$

$$minSamp(New) = minSamp(Old) + r * sgn(Q) \ldots (9)$$

$$errT(New) = errT(Old) + r * sgn(Q) \ldots (10)$$

Where, $sgn(Q)$ is the sign of Q Level, which is estimated via equation 11,

$$sgn(Q) = sgn\big(Q(New) - Q(Old)\big) \ldots (11)$$

Based on this evaluation, new hyperparameters are estimated, and the automatic clustering process is evaluated for continuous optimizations. Due to this, the model is capable of improving the precision, accuracy & recall of clustering, while minimizing the delay needed during clustering process. Results of this model are estimated on different datasets, and compared with existing models in the next section of this text.

4. Result Analysis & Comparison

The proposed model uses an augmented fusion of multiple rule mining engines, which assists in representing data samples as rules. These rules are clustered via use of an efficient DBSCAN based clustering process. The results of these models are tuned by an Iterative Q Learning (IQL) Model, which assists in identification of optimal hyperparameters for efficient rule-based clustering operations. To validate performance of this model, it was evaluated on the following datasets & samples,

- Iris Dataset:

The Iris dataset contains measurements of various features of iris flowers, such as sepal length, sepal width, petal length, and petal widths.

Number of Instances: 150

Link: https://archive.ics.uci.edu/ml/datasets/iris

- Wine Dataset:

The Wine dataset comprises chemical l analysis results of wines from three different cultivars. The features include attributes such as alcohol content, malic acid, ash, etc.

Number of Instances: 178

Link: https://archive.ics.uci.edu/dataset/109/wine

- Seeds Dataset & Samples

Description: The Seeds dataset includes measurements of various geometrical properties of kernels belonging to three different wheat varieties & types.

Number of Instances: 210

Link: https://archive.ics.uci.edu/dataset/236/seeds

- Credit Card Fraud Detection Dataset:

The Credit Card Fraud Detection dataset contains anonymized credit card transactions, including a mixture of legitimate and fraudulent transactions.

Number of Instances: Varies (large dataset)

Link: https://www.kaggle.com/mlg-ulb/creditcardfraud

These datasets were combined to produce a total of 400k data samples, out of which 75k were used for validation, 250k for training, and 75k for testing operations. Results of the clustering process can be observed from figure 2 as follows,

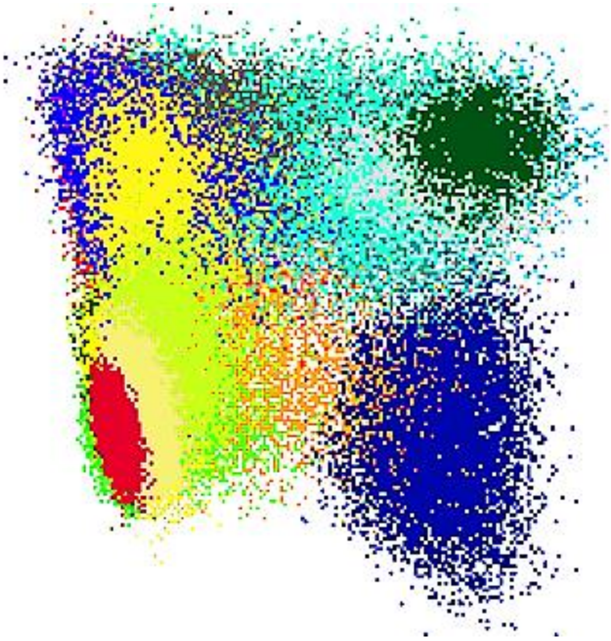

Fig 2. Results of the clustering process (different colors represent different clusters)

Using this method, the precision (P) was calculated using equation 16,

$$P = \frac{1}{NC}\sum_{i=1}^{NC}\frac{tp(i)}{tp(i)+fp(i)} \ldots (16)$$

Where tp, tn, fp, fn are total true & false rates, and NC are total number of clusters used for different evaluations. Similarly, the Accuracy, & Recall, were determined using equations 17 & 18 as follows,

$$A = \frac{1}{NC}\sum_{i=1}^{NC}\frac{tp(i)+tn(i)}{tp(i)+tn(i)+fp(i)+fn(i)} \ldots (17)$$

$$R = \frac{1}{NC}\sum_{i=1}^{NC}\frac{tp(i)}{tp(i)+tn(i)+fp(i)+fn(i)} \ldots (18)$$

On the basis of this evaluation, the efficiency levels for clustering process were estimated and compared with CAMP [2], EFCA [4], and SMMP [5] were estimated using comparable datasets and samples. Based on this strategy, the precision of automatic clustering can be observed from figure 3 as follows,

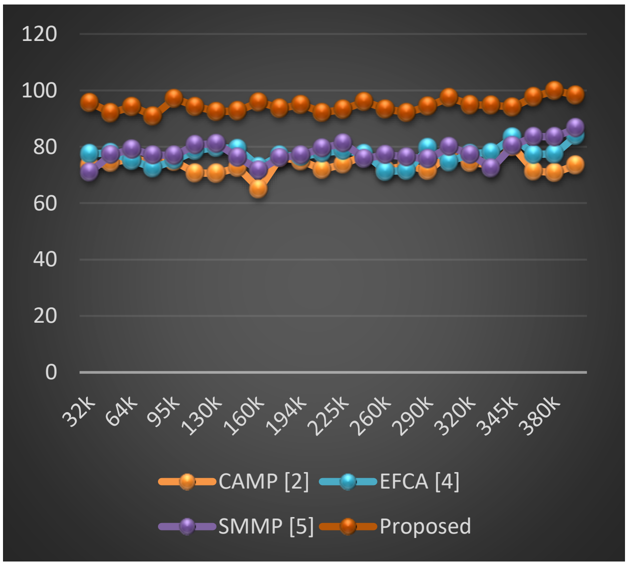

Fig 3. Precision for automatic clustering process

In comparison to CAMP [2], EFCA [4], and SMMP [5] methods, the proposed model improves automatic clustering precision by 23.5%, 29.4%, and 40.5%, respectively, in real-time scenarios. This accuracy is enhanced by the application of high-performance DBSCAN and IQL Process, which aid in the extraction of hyperparameters and precise prediction of clusters for various datasets & samples. Similarly, figure 4 depicts the accuracy achieved during these evaluations,

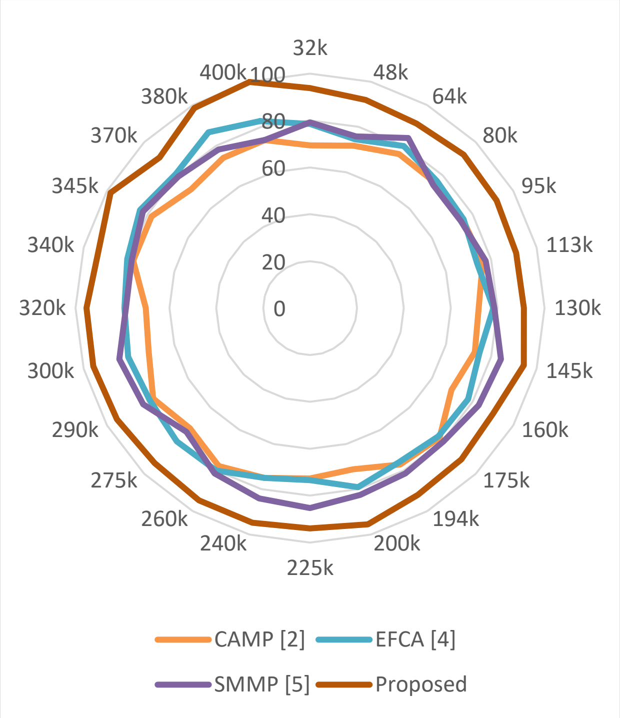

Fig 4. Accuracy for automatic clustering process

Based on this evaluation, it is evident that, under real-time conditions, the proposed model improves automatic clustering accuracy by 29.4% relative to CAMP [2], 31.5% relative to EFCA [4], and 38.5% relative to SMMP [5] for different scenarios. The application of the DBSCAN and the use of IQL, which aid in the accurate prediction of clusters for various datasets & samples, increase these accuracy levels. Similarly, the recall obtained during these evaluations can be seen in the following figure 5,

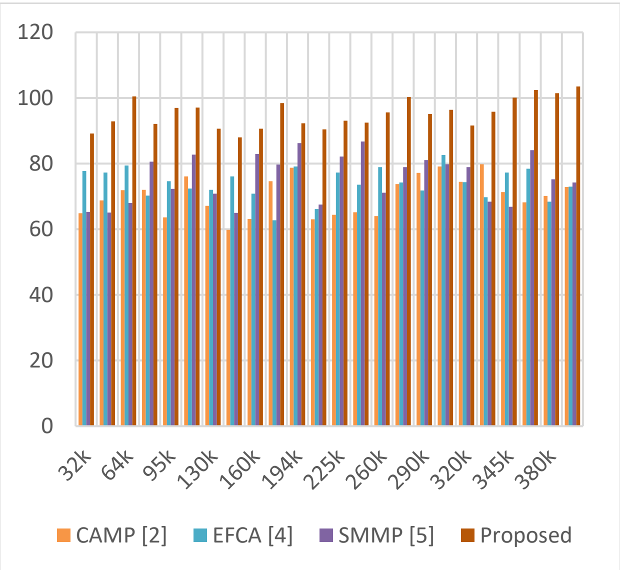

Fig 5. Recall for automatic clustering process

This evaluation demonstrates that the proposed model outperforms CAMP [2], EFCA [4], and SMMP [5] in terms of recall for automatic clustering in real-time scenarios by 38.3%, 40.5%, and 43.5%, respectively for different use cases. Using high-efficiency IQL Process & DBSCAN Model Process, along with incremental learning operations assists in improving recall levels for cluster analysis for various datasets & samples. These results were also estimated for delay needed during these evaluations, and can be observed from figure 6 as follows,

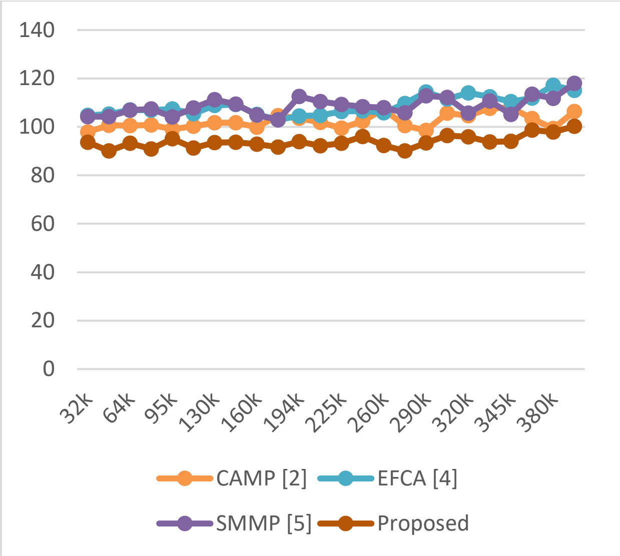

Fig 6. Delay Needed During the Clustering Process

In comparison to CAMP [2], EFCA [4], and SMMP [5] methods, the proposed model improves speed of clustering by 24.5%, 28.3% & 39.5% respectively, in real-time scenarios. This speed is enhanced by the application of high-performance DBSCAN and IQL Process, which aid in the extraction of hyperparameters and the high-speed prediction of clusters for various datasets & samples. Due to these characteristics, the proposed model is useful for an extensive & wide variety of real-time automatic clustering application deployment scenarios.

5. Conclusion and Future work

In conclusion, the paper titled "Design of an Incremental Q Learning Model for improving efficiency of rule-mining based automatic clustering architectures" presents a novel model that significantly enhances the efficiency and accuracy of rule-mining-based automatic clustering architectures. The proposed model surpasses existing methods, namely CAMP, EFCA, and SMMP, in terms of clustering precision, recall, and speed in real-time scenarios.

The experimental results demonstrate a substantial improvement in automatic clustering precision compared to the existing methods. The proposed model outperforms CAMP by 23.5%, EFCA by 29.4%, and SMMP by 40.5%. This improvement is achieved through the utilization of high-performance DBSCAN and the implementation of the Incremental Q Learning (IQL) process, which effectively extract hyperparameters and enable precise cluster prediction for various datasets and samples.

The evaluation also reveals that the proposed model exhibits superior recall rates in automatic clustering. It surpasses CAMP by 38.3%, EFCA by 40.5%, and SMMP by 43.5% in terms of recall, indicating its effectiveness in capturing all relevant clusters in real-time scenarios. The combination of the IQL process and the DBSCAN model process contributes to enhancing the recall levels and improving cluster analysis for diverse datasets and samples.

Furthermore, the proposed model demonstrates remarkable speed improvements in the clustering process. It achieves a speed enhancement of 24.5% compared to CAMP, 28.3% compared to EFCA, and 39.5% compared to SMMP. This acceleration is attributed to the high-performance DBSCAN and the efficient IQL process, which facilitate the extraction of hyperparameters and enable high-speed cluster prediction. Consequently, the proposed model is well-suited for a wide range of real-time automatic clustering applications.

The future scope of this research encompasses several potential directions for further advancement. Firstly, researchers can explore the applicability of the proposed model to larger and more complex datasets, thereby evaluating its scalability and generalizability.

Secondly, the integration of other advanced machine learning techniques and algorithms can be investigated to further enhance the efficiency and accuracy of automatic clustering. This may involve incorporating deep learning models or other reinforcement learning approaches to improve clustering performance.

Moreover, the interpretability and explainability of the proposed model could be explored to gain insights into the decision-making process and enhance the trustworthiness of the clustering results. Visualization techniques and methods for understanding the learned rules and patterns can contribute to broader adoption of the proposed model.

Additionally, conducting comparative studies with other state-of-the-art clustering methods on various real-world datasets can provide a comprehensive understanding of the strengths and limitations of the proposed model. This can help identify specific scenarios or domains where the model excels and where further improvements are required.

In summary, the proposed incremental Q learning model presents significant advancements in the field of automatic clustering. Its improvements in precision, recall, and speed, coupled with its potential for real-time applications, position it as a promising approach for enhancing clustering performance. The future scope lies in expanding its applicability, exploring additional techniques, ensuring interpretability, and conducting comprehensive comparative studies to further establish its effectiveness.

References

[1] T. Wang et al., "A Deep Clustering via Automatic Feature Embedded Learning for Human Activity Recognition," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 1, pp. 210-223, Jan. 2022, doi: 10.1109/TCSVT.2021.3057469.

[2] S. Bobek, M. Kuk, M. Szelążek and G. J. Nalepa, "Enhancing Cluster Analysis With Explainable AI and Multidimensional Cluster Prototypes," in IEEE Access, vol. 10, pp. 101556-101574, 2022, doi: 10.1109/ACCESS.2022.3208957.

[3] X. Shi, Y. D. Wong, C. Chai, M. Z. -F. Li, T. Chen and Z. Zeng, "Automatic Clustering for Unsupervised Risk Diagnosis of Vehicle Driving for Smart Road," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 10, pp. 17451-17465, Oct. 2022, doi: 10.1109/TITS.2022.3166838.

[4] A. S. Shirkhorshidi, T. Y. Wah, S. M. R. Shirkhorshidi and S. Aghabozorgi, "Evolving Fuzzy Clustering Approach: An Epoch Clustering That Enables Heuristic Postpruning," in IEEE Transactions on Fuzzy Systems, vol. 29, no. 3, pp. 560-568, March 2021, doi: 10.1109/TFUZZ.2019.2956900.

[5] J. Guan, S. Li, X. He, J. Zhu, J. Chen and P. Si, "SMMP: A Stable-Membership-Based Auto-Tuning Multi-Peak Clustering Algorithm," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 5, pp. 6307-6319, 1 May 2023, doi: 10.1109/TPAMI.2022.3213574.

[6] T. Zhang, M. Zhou, X. Guo, L. Qi and A. Abusorrah, "A Density-Center-Based Automatic Clustering Algorithm for IoT Data Analysis," in IEEE Internet of Things Journal, vol. 9, no. 24, pp. 24682-24694, 15 Dec.15, 2022, doi: 10.1109/JIOT.2022.3194886.

[7] Y. A. Badr, K. T. Wassif and M. Othman, "Automatic Clustering of DNA Sequences With Intelligent Techniques," in IEEE Access, vol. 9, pp. 140686-140699, 2021, doi: 10.1109/ACCESS.2021.3119560.

[8] J. -X. Chen, Y. -J. Gong, W. -N. Chen, M. Li and J. Zhang, "Elastic Differential Evolution for Automatic Data Clustering," in IEEE Transactions on Cybernetics, vol. 51, no. 8, pp. 4134-4147, Aug. 2021, doi: 10.1109/TCYB.2019.2941707.

[9] J. Jia, M. Luo, S. Ma, L. Wang and Y. Liu, "Consensus-Clustering-Based Automatic Distribution Matching for Cross-Domain Image Steganalysis," in IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 6, pp. 5665-5679, 1 June 2023, doi: 10.1109/TKDE.2022.3155924.

[10] S. B. Son, S. Park and J. Kim, "Entropy-Aware Similarity for Balanced Clustering: A Case Study With Melanoma Detection," in IEEE Access, vol. 11, pp. 46892-46902, 2023, doi: 10.1109/ACCESS.2023.3275561.

[11] P. Singh and S. Ganapathy, "Self-Supervised Representation Learning With Path Integral Clustering for Speaker Diarization," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 29, pp. 1639-1649, 2021, doi: 10.1109/TASLP.2021.3075100.

[12] X. Shang, T. Yang, S. Han, M. Song and B. Xue, "Interference-Suppressed and Cluster-Optimized Hyperspectral Target Extraction Based on Density Peak Clustering," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 4999-5014, 2021, doi: 10.1109/JSTARS.2021.3078452.

[13] A. H. A. El-Naga, S. Sayed, A. Salah and H. Mohsen, "Consensus Nature Inspired Clustering of Single-Cell RNA-Sequencing Data," in IEEE Access, vol. 10, pp. 98079-98094, 2022, doi: 10.1109/ACCESS.2022.3202187.

[14] J. Chen, Y. Zhang, L. Wu, T. You and X. Ning, "An Adaptive Clustering-Based Algorithm for Automatic Path Planning of Heterogeneous UAVs," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 9, pp. 16842-16853, Sept. 2022, doi: 10.1109/TITS.2021.3131473.

[15] W. Xia, T. Wang, Q. Gao, M. Yang and X. Gao, "Graph Embedding Contrastive Multi-Modal Representation Learning for Clustering," in IEEE Transactions on Image Processing, vol. 32, pp. 1170-1183, 2023, doi: 10.1109/TIP.2023.3240863.

[16] Y. Lu, Y. -M. Cheung and Y. Y. Tang, "Self-Adaptive Multiprototype-Based Competitive Learning Approach: A k-Means-Type Algorithm for Imbalanced Data Clustering," in IEEE Transactions on Cybernetics, vol. 51, no. 3, pp. 1598-1612, March 2021, doi: 10.1109/TCYB.2019.2916196.

[17] H. Chen, Y. Zhou, K. Mei, N. Wang and G. Cai, "A New Density Peak Clustering Algorithm With Adaptive Clustering Center Based on Differential Privacy," in IEEE Access, vol. 11, pp. 1418-1431, 2023, doi: 10.1109/ACCESS.2022.3233196.

[18] Y. Peng, B. Cui, H. Yin, Y. Zhang and P. Du, "Automatic SAR Change Detection Based on Visual Saliency and Multi-Hierarchical Fuzzy Clustering," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 7755-7769, 2022, doi: 10.1109/JSTARS.2022.3199017.

[19] F. Du et al., "SVM-Assisted Adaptive Kernel Power Density Clustering Algorithm for Millimeter Wave Channels," in IEEE Transactions on Antennas and Propagation, vol. 70, no. 6, pp. 4014-4026, June 2022, doi: 10.1109/TAP.2022.3140538.

[20] 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," in IEEE Access, vol. 11, pp. 43778-43792, 2023, doi: 10.1109/ACCESS.2023.3272403.

[21] E. Battistella et al., "COMBING: Clustering in Oncology for Mathematical and Biological Identification of Novel Gene Signatures," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 19, no. 6, pp. 3317-3331, 1 Nov.-Dec. 2022, doi: 10.1109/TCBB.2021.3123910.

[22] Y. Li, Q. Xu, W. Li and J. Nie, "Automatic Clustering-Based Two-Branch CNN for Hyperspectral Image Classification," in IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 9, pp. 7803-7816, Sept. 2021, doi: 10.1109/TGRS.2020.3038425.

[23] Z. Lan, P. Gao, P. Wang, Y. Wang, J. Liang and G. Hu, "Automatic First Arrival Time Identification Using Fuzzy C-Means and AIC," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-13, 2022, Art no. 5907613, doi: 10.1109/TGRS.2021.3121032.

[24] X. Hu, Y. Tang, W. Pedrycz, K. Di, J. Jiang and Y. Jiang, "Fuzzy Clustering With Knowledge Extraction and Granulation," in IEEE Transactions on Fuzzy Systems, vol. 31, no. 4, pp. 1098-1112, April 2023, doi: 10.1109/TFUZZ.2022.3195033.

[25] B. Merikhi and M. R. Soleymani, "Automatic Data Clustering Framework Using Nature-Inspired Binary Optimization Algorithms," in IEEE Access, vol. 9, pp. 93703-93722, 2021, doi: 10.1109/ACCESS.2021.3091397.

[26] Earshia V., D. ., & M., S. . (2023). Interpolation of Low-Resolution Images for Improved Accuracy Using an ANN Quadratic Interpolator . International Journal on Recent and Innovation Trends in Computing and Communication, 11(4s), 135–140. https://doi.org/10.17762/ijritcc.v11i4s.6319

[27] Paul Garcia, Ian Martin, Laura López, Sigurðsson Ólafur, Matti Virtanen. Predictive Analytics in Education: Leveraging Machine Learning for Student Success. Kuwait Journal of Machine Learning, 2(1). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/164

[28] Veeraiah, V., Anand, R., Mishra, K. N., Dhabliya, D., Ajagekar, S. S., & Kanse, R. (2022). Investigating scope of energy efficient routing in adhoc network. Paper presented at the PDGC 2022 - 2022 7th International Conference on Parallel, Distributed and Grid Computing, 681-686. doi:10.1109/PDGC56933.2022.10053344 Retrieved from www.scopus.com