International Journal of Advanced Computer Technology
ISSN: 2319-7900 www.ijact.org Volume 12, Issue 4, July-August 2023

A Review on K-means Clustering Based on Quantum Particle Swarm Optimisation Algorithm

Shilpa Meshram, Jayshree Boaddh
Department of CSE, Mittal Institute of Technology, Bhopal, India
shilpameshram028@gmail.com, Jayshree.boaddh@gmail.com

Abstract:- Unsupervised learning clustering techniques play a vital role in data mining, with a wide range of applications in unsupervised classification. Clustering is a method used to categorise data into meaningful groups. The k-means algorithm is a well-known clustering algorithm that aims to minimise the squared distance between feature values of points within the same cluster. In many applications, using an evolutionary computation technique called Quantum Particle Swarm Optimization (QPSO) in conjunction with the k-means algorithm has proven effective in finding suboptimal solutions. In this algorithm, the cluster centres are simulated as particles, allowing for the identification of suitable and stable cluster centres. This paper discusses the current improvement in the QPSO-k-means clustering algorithm, focusing on swarm initialisation and algorithm parameter optimisation. We validate the algorithm using the UCI healthcare dataset and demonstrate its ability to address suboptimal clustering by optimising parameters such as the number of iterations, error rate, and optimal solution for cluster centres. The minimisation factor of the validation parameter indicates the compactness and validity of the clustering algorithm.

Keywords: Data Mining, Unsupervised Learning, Clustering, QPSO-K-Means Clustering Algorithm

I. INTRODUCTION

Data mining has gained significant attention in the information industry due to the abundance of large datasets and the need to transform data into valuable information and knowledge. In today's highly competitive market, timely and accurate information is crucial in decision-making. With vast data available in the real world, extracting useful information from such a massive database within the required timeframe and in the desired pattern poses a significant challenge. Recent advancements in information technology have led to the generation of vast quantities of data, often unstructured and challenging to analyse. Data clustering has emerged as a widely used technique to gain insights, summarise data, identify natural patterns, and discover hidden patterns. Clustering, also known as clustering analysis, involves partitioning a dataset into meaningful groups (clusters), where objects within the same group are similar to each other and dissimilar to objects in other groups. It addresses the unsupervised classification problem, where classes are not known in advance [1]. Clustering is a vital technique in data mining and finds applications in various domains, including machine learning. It enables the division of data objects into groups based on available information, such as object descriptions and relationships among them. This feature has widespread applicability, including knowledge discovery, vector quantisation, pattern recognition, data mining, and data dredging. The main objective of clustering is to segment a large dataset into meaningful clusters. Two commonly used methods for clustering are hierarchical clustering and partitioned clustering. While effective in generating a local suboptimal solution, the K-means algorithm is limited in its scope. In contrast, the Quantum Particle Swarm Optimization based on the K-means clustering algorithm (QPSO-KMCA) offers a globalised search methodology that can be incorporated into the K-means algorithm to obtain a global suboptimal solution. By leveraging the advantages of both algorithms, it is possible to overcome their drawbacks

and develop a combined algorithm for more effective solutions [2].

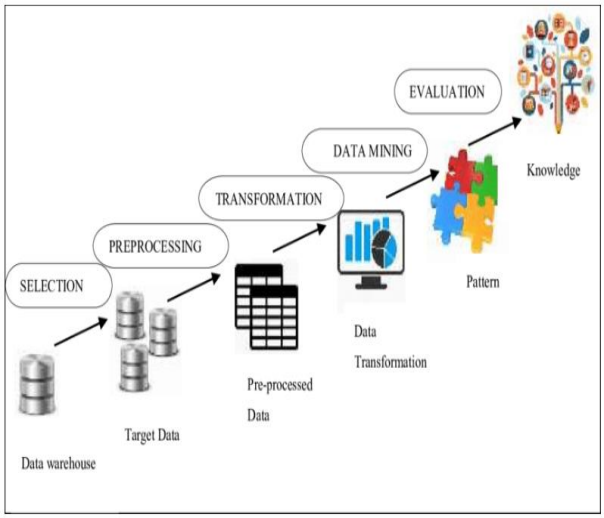

Figure 1. Knowledge Discovery Database (KDD) Process

1.1 Clustering

It is a type of unsupervised learning method. An unsupervised learning method is a method in which we draw references from datasets consisting of input data without labelled responses. Generally, it is used to find meaningful structure, underlying explanatory processes, generative features, and groupings inherent in a set of examples.

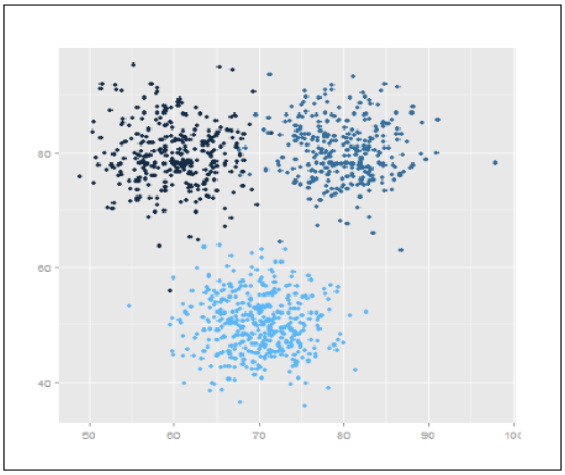

Figure 2. Clustering process in different clusters

Clustering is dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to those in other groups. It is a collection of objects based on similarities and dissimilarities between them. For example, the data points in the graph below clustered together can be classified into one single group. We can distinguish the clusters, and we can identify that there are 3 clusters in the below picture [3].

1.2 Quantum Particle Swarm Optimisation Based on K-means Clustering Algorithm (QPSOKMCA)

Initially proposed by M QPSOKMCA, the K-means algorithm is widely used in knowledge discovery and data mining to solve cluster analysis problems. However, further research has revealed several issues with the K-means algorithm. As society progresses, the types and quantity of data are increasing rapidly. With the advent of computer information networks and low-cost data storage, generating large amounts of data has become effortless, leading to the growth of social data. While the K-means algorithm offers advantages such as fast operation speed and a simple process, it is also sensitive to the initial values and prone to local optima, limiting its effectiveness and hindering its broader application [4]. In order to address these limitations, the Quantum Particle Swarm Optimization based on K-means Clustering Algorithm (QPSOKMCA) was developed as a new algorithm. QPSOKMCA is a simplified algorithm based on the genetic algorithm, eliminating the “selection” and “variation” processes and utilising group optimal solutions to determine the global optimal solution [5]. This bionic optimisation algorithm achieves high accuracy, fast operation speed, and simple steps, making it widely applicable across various fields. Through extensive research and analysis, it has been observed that QPSOKMCA exhibits strong capabilities in global search and local optimisation, effectively compensating for the shortcomings of the K-means algorithm. By leveraging the strengths of both the K-means algorithm and the QPSOKMCA, researchers aim to improve the performance and expand the application scope of clustering analysis in fields such as image segmentation, feature extraction, data analysis, and pattern recognition [6].

II. RELATED WORK

Several studies have been conducted to address the challenges and improve the performance of clustering algorithms. Here are some relevant works: Lili Bai et al. [7] proposed a k-means clustering algorithm based on an improved quantum particle swarm optimisation algorithm. This approach simulates the cluster centre as a particle and utilises cloning and mutation operations to enhance diversity and global search ability. The algorithm achieves more accurate clustering results and ensures global convergence. Semeh Ben et al. [8] discussed the limitations of the k-means algorithm for categorical clustering and introduced the k-modes algorithm as an extension to handle categorical datasets. They also presented a new categorical method called MFk-M, which converts initial categorical data into numeric values based on relative frequency. RSM Lakshmi et al. [9] highlighted the significance of clustering algorithms in generating insights from large volumes of data. They emphasised the importance of proper classification and defining properties for clustering algorithms, particularly when dealing with big data. Rezaee Jordehi et al. [8] examined various strategies adopted in particle swarm optimisation (PSO) for solving discrete optimisation problems. They analysed the pros and cons of each strategy and its applicability in tackling discrete problems. Poli et al. [11] comprehensively analysed PSO applications and categorised many publications stored in the IEEE Xplore database. Their work aimed to provide an up-to-date overview of the achievements and advancements in the field of PSO applications. Kohei Arai et al. [12] proposed a new approach to optimise the initial centroids for the k-means algorithm. They utilised multiple clustering results of k-means, including those reaching local optima, and combined them with the hierarchical algorithm to determine improved initial centroids.

Chouhan et al. (2018) [16] introduced a method combining PSO and k-means for document clustering. PSO was used to locate the best points in the search space, which were then used as initial cluster centroids in the k-means algorithm to improve clustering results. Janani et al. (2019) [17] developed a spectral clustering algorithm with PSO (SCPSO) for text document clustering. They utilised global and local optimisation functions to randomise the initial population and combined spectral clustering with swarm optimisation to handle large amounts of text documents. Barakbah et al. [12] proposed a Centronit algorithm to optimise the initial centroids for k-means clustering. Their approach calculates the average distance of the nearest data inside the minimum distance region to determine the initial centroids, effectively improving clustering results and robustness against outliers. Caron, Mathilde et al. [18] mentioned that clustering has been widely applied and studied in computer vision. Still, little work has been done to adapt it for end-to-end training of visual features on large-scale datasets.

III. EXPECT THE OUTCOME

The expected outcome of using data mining based on quantum particle swarm optimisation with k-means clustering (QPSOKMCA) is to achieve improved clustering results and a global suboptimal solution with a low error rate. The QPSOKMCA technique offers a globalised search methodology, which helps overcome the local suboptimal solutions generated by the traditional K-means algorithm. By combining the advantages of both algorithms, our proposed approach aims to provide a more accurate and effective solution for data analysis. The primary objective of our study is to apply the QPSOKMCA algorithm to analyse a healthcare dataset. By utilising the algorithm's global search capabilities, we expect to obtain a clustering outcome that minimises the error rate and ensures the best possible solution. It is crucial in healthcare data analysis, where accurate classification and identification of patterns are essential for decision-making processes. The expected outcome of our research is twofold. Firstly, we anticipate that the QPSOKMCA algorithm will significantly improve the clustering results compared to traditional K-means clustering. The algorithm's ability to find global suboptimal solutions will enhance the accuracy and precision of the clusters formed, leading to more meaningful insights and patterns within the healthcare

dataset. Secondly, we aim to determine the lowest possible error rate for the healthcare dataset analysis. By optimising the algorithm's parameters and conducting rigorous experimentation, we seek to achieve a clustering outcome that minimises errors and maximises the accuracy of the clustering process. This outcome will contribute to the reliability and effectiveness of the data mining results, enabling better decision-making in healthcare applications. Overall, we expect data mining based on QPSOKMCA will provide superior clustering results, reduce errors, and deliver valuable insights from the healthcare dataset, ultimately enhancing the understanding and utilisation of the data for improved decision-making processes.

CONCLUSION

this study addressed the limitations of the K-means clustering algorithm by introducing a modified version based on quantum-behaved particle swarm optimisation (QPSO). The K-means algorithm was observed to be sensitive to initial conditions, leading to convergence on suboptimal solutions. To overcome this, the QPSO was integrated with the K-means algorithm to improve its performance. The proposed algorithm successfully solved data mapping challenges and objective function minimisation through particle swarm optimisation. By incorporating the QPSO, the algorithm demonstrated enhanced time complexity and reduced computational errors, particularly when applied to standard datasets. However, it was noted that QPSOs based on the K-means algorithm often converge quickly but are prone to local optima due to imbalances in global and local search capabilities. Several improvement algorithms, such as the QPSO algorithm based on data analysis models and adaptive multi-objective QPSO algorithms based on K-means clustering, have been proposed to address this issue. Further research is required to explore the search process for obtaining suboptimal solutions and determining the optimal balance between global and local search capabilities. Additionally, the proposed algorithm was evaluated on healthcare and real-world datasets, demonstrating its effectiveness in achieving accurate clustering and minimising errors. In summary, the developed algorithm shows promise in improving the performance of K-means clustering for data mining applications, particularly in healthcare dataset analysis. It offers a potential solution to finding the lowest possible error in clustering and can be applied to various real-world datasets. Continued research and development in this area will contribute to further data mining and clustering techniques advancements.

REFERENCES

[1]. Jayamalini, K., and M. Ponnavaikko. "Research on web data mining concepts, techniques and applications." In 2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET), pp. 1-5. IEEE, 2017.

[2]. Gera, Mansi, and Shivani Goel. "Data mining-techniques, methods and algorithms: A review on tools and their validity." International Journal of Computer Applications 113, no. 18, 2015.

[3]. Sisodia, Deepti, Lokesh Singh, Sheetal Sisodia, and Khushboo Saxena. "Clustering techniques: a brief survey of different clustering algorithms." International Journal of Latest Trends in Engineering and Technology (IJLTET) 1, no. 3: 82-87, 2012.

[4]. Zou, Hailei. "Clustering algorithm and its application in data mining." Wireless Personal Communications 110, no. 1: 21-30, 2020.

[5]. Mythili, S., and E. Madhiya. "An analysis on clustering algorithms in data mining." International Journal of Computer Science and Mobile Computing 3, no. 1 334-340, 2014.

[6]. J. Liu, L. Han and L. Hou, "K-Mean Clustering Algorithm Based on Particle Swarm Optimisation," System Engineering Theory and Practice, vol. 06, pp. 54-58, 2005.

[7]. Bai, Lili, Zerui Song, Haijie Bao, and Jingqing Jiang. "K-Means clustering based on improved quantum particle swarm optimisation algorithm." In 2021 13th International Conference on Advanced Computational Intelligence (ICACI), pp. 140-145. IEEE, 2021.

[8]. Salem, Semeh Ben, Sami Naouali, and ZiedChtourou. "A fast and effective partitional clustering algorithm for large categorical datasets using a k-means based approach." Computers & Electrical Engineering 68, 463-483, 2018.

[9]. Patibandla, RSM Lakshmi, and N. Veeranjaneyulu. "Survey on clustering algorithms for unstructured data." In Intelligent Engineering Informatics, pp. 421-429. Springer, Singapore, 2018.

[10]. Rezaee Jordehi, Ahmad, and Jasronita Jasni. "Particle swarm optimisation for discrete optimisation problems: a review." Artificial Intelligence Review 43: 243-25, 2015.

[11]. Poli, Riccardo. "Analysis of the publications on the applications of particle swarm optimisation." Journal of Artificial Evolution and Applications 2008: 1-10, 2008.

[12]. Arai, Kohei, and Ali Ridho. "Hierarchical K-means: an algorithm for centroids initialisation for K-means.", Reports of the Faculty of Science and Engineering, Vol. 36, No.1, 2007.

[13]. R. Chouhan and A. Purohit, "An approach for document clustering using PSO and K-means algorithm," in 2018 2nd International Conference on Inventive Systems and Control (ICISC), IEEE, pp. 1380-1384, 2018.

[14]. R. Janani and S. Vijayarani, "Text document clustering using spectral clustering algorithm with particle swarm optimisation," Expert Systems with Applications, vol. 134, pp. 192-200, 2019.

[15]. Barakbah, Ali Ridho, and Kohei Arai. "Centronit: Initial Centroid Designation Algorithm for K-Means Clustering." EMITTER International journal of engineering technology 2, no. 1: 50-62, 2014.

[16]. Caron, Mathilde, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. "Deep clustering for unsupervised learning of visual features." In Proceedings of the European Conference on Computer Vision (ECCV), pp. 132-149. 2018.