Characteristics-Based Least Common Multiple: A Novel Clustering Algorithm to Optimize Indoor Positioning

Hamaad Rafique[a], Davide Patti[b], Maurizio Palesi[c] and Gaetano Carmelo La Delfa
Department of Electrical, Electronics and Computer Engineering, University of Catania, Catania, Italy
hamaad.rafique@phd.unict.it, {davide.patti, maurizio.palesi, gaetano.ladelfa}@unict.it

Keywords: Indoor Localization, Novel Clustering Technique, Least Common Multiple (LCM), Machine Learning, Clustering.

Abstract: Clustering is an unsupervised learning technique for grouping data based on similarity criteria. Conventional clustering algorithms like K-Means and agglomerative clustering often require predefined parameters such as the number of clusters and struggle to identify irregularly shaped clusters. Additionally, these methods fail to correctly cluster magnetic field signals with similar characteristics used for positioning in magnetic fingerprint-based indoor localization. This paper introduces a novel Characteristics-Based Least Common Multiple (LCM) clustering algorithm to address these limitations. This algorithm autonomously determines the number and shape of clusters and correctly classifies misclassified points based on characteristic similarities using LCM-based techniques. The effectiveness of the proposed technique was evaluated using state-of-the-art metrics like the Silhouette score, Calinski-Harabasz Index, and Davies-Bouldin Index on benchmark datasets.

1 INTRODUCTION

The advancement in IoT technology has led to a rise in data-intensive applications like indoor localization within various industries. Indoor localization aims to determine the user's location within an indoor environment, where GPS signals are ineffective due to thick building structures or basements. Researchers are developing alternative indoor localization systems (ILS) using sensor data from WiFi, RFID, Bluetooth Low Energy, and Magnetic Field signals (MFS) (Rafique et al., 2023a). However, factors like perturbations and ferromagnetic materials introduce complexities in MFS, causing inaccuracies in location predictions. Machine learning, specifically unsupervised clustering techniques such as K-means (Berahmand et al., 2022), density-based methods like k-mean kernel (Anuwatkun et al., 2019), DBSCAN (Ester et al., 1996), and OPTICS (Ankerst et al., 1999), has been used to manage large datasets but faces challenges in performing optimally on substantial localization datasets. This leads to misclassifying data points that are far apart but have similar features, causing multiple predictions for different locations.

To address these challenges, a new clustering technique called the characteristic-based **L**east **C**ommon **M**ultiple (LCM) technique is proposed. This method focuses on clustering based on similar characteristics, addressing the limitations of conventional clustering methods, such as reliance on user-defined parameters and difficulties with arbitrary cluster shapes. It efficiently handles the issue of misplaced data points within clusters caused by the overlapping behaviour of MFS due to ferromagnetic materials in indoor environments. This technique aims to enhance indoor location tracking and reliable data clustering, which are important for applications like asset tracking, user navigation, and context-aware services. The research question addressed can be defined as "How can we accurately identify and correctly classify data points or sub-clusters that share similar characteristics but are physically separated and thus placed in incorrect clusters?". Figure 1 presents the conceptual representation of the research question. The critical contributions of this work can be summarised as follows:

- We introduce the "Characteristics-Based Least Common Multiple Clustering Algorithm" to perform arbitrary shape clustering for enhancing indoor localization.
- Our novel method uses LCM to reveal the unique properties of magnetic field sensors.

[a] https://orcid.org/0000-0003-4272-5360
[b] https://orcid.org/0000-0003-0874-7793
[c] https://orcid.org/0000-0003-3129-0664

Rafique, H., Patti, D., Palesi, M. and La Delfa, G.
Characteristics-Based Least Common Multiple: A Novel Clustering Algorithm to Optimize Indoor Positioning.
DOI: 10.5220/0012943900003822

In *Proceedings of the 21st International Conference on Informatics in Control, Automation and Robotics (ICINCO 2024) - Volume 1*, pages 301-308
ISBN: 978-989-758-717-7; ISSN: 2184-2809

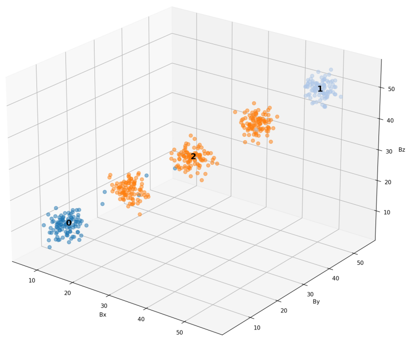


Figure 1: Conceptual illustration of misclassified data points with similar characteristics that are assigned to cluster two. Dark blue data points in cluster two belong to cluster zero. Likewise, a light blue data point belongs to cluster one. This miss-classification leads to incorrect positioning. B_x, B_y, and B_z correspond to the three dimensions of MFS.

- It effectively organizes samples into distinct sub-clusters, correcting cases where they are wrongly grouped with similar characteristics.
- We introduced a tunable parameter to improve clusters.
- We tested our technique with advanced metrics like the Silhouette Score, Calinski-Harabasz Index, and Davies-Bouldin Index on benchmark datasets.

The paper is structured as follows: Section 1 provides a brief introduction. Section 2 discusses the research motivation and supporting studies. The proposed clustering technique is detailed in Section 3. Section 4 covers the state-of-the-art evaluation techniques and the data used for evaluation. The results obtained using these metrics are presented in Section 5. Finally, the conclusion is provided in Section 6.

2 RELATED WORK

In the IoT era, technological advancements have significantly increased data-intensive applications across various industries, generating vast amounts of data in healthcare, transportation, agriculture, smart cities, security, and localization. Clustering techniques have been widely proposed to understand better and organize this data. Clustering, a fundamental challenge in data mining, involves organizing datasets into distinct groups based on their similarities (Xu et al., 2024; Fang et al., 2023; Vinciguerra et al., 2024; Rafique et al., 2020). Various algorithms, classified as hierarchical and partitional, have been developed, with partitional clustering techniques like k-means being popular for their efficiency and simplicity (Berahmand et al., 2022; Ren et al., 2019). However, k-means has limitations, including reliance on a user-defined number of clusters and only producing spherical clusters (Junyi et al., 2021; Lee and Lee, 2020). Other methods like PF Clust (Mavridis et al., 2013) and automated clustering with force functions aim to address these issues but still struggle with arbitrary-shaped clusters and time-consuming processes (Vo-Van et al., 2020; El Khediri et al., 2020; Rafique et al., 2023c; Von Luxburg, 2007; Singh and Soni, 2019; Rafique et al., 2023b) .

Density-based clustering techniques, such as DBSCAN, have emerged as effective solutions for identifying arbitrary-shaped clusters without user-defined parameters (Ester et al., 1996). DBSCAN uses a density threshold to cluster data points, handling arbitrary shapes well but sometimes merging close clusters. Variants like (Anuwatkun et al., 2019; Ankerst et al., 1999; Junyi et al., 2021; Vo-Van et al., 2020) address this by ordering data points based on density to reveal clusters of different densities. Additionally, deep learning-based clustering techniques have integrated deep learning with traditional methods to handle complex patterns in high-dimensional spaces, proving effective in tasks like image segmentation and document clustering. Techniques like unified and discrete bipartite graph learning and strong augmentation clustering have shown robustness and efficiency in multi-view datasets (Fang et al., 2023). Time series-based clustering techniques, such as k-shape (Paparrizos and Gravano, 2015; Cui et al., 2021), have also been proposed to address the unique properties of time series data, offering robust performance across various metrics.

Despite these advancements, existing clustering methods often overlook scenarios where data points with similar characteristics are physically distant, leading to incorrect clustering within the localization domain. This issue is critical in applications like indoor localization, where high accuracy is essential for effective resource management, security, safety, and navigation. To address these limitations, we propose the ”Characteristic-Based Least Common Multiple Technique.” This new algorithm leverages MFS to detect physically distant sample points with similar characteristics and appropriately assigns them to their respective clusters, improving the accuracy and effectiveness of indoor positioning.

3 PROPOSED ALGORITHM

Unlike traditional methods that initiate clustering from a central point or the densest area, this technique emphasizes the unique features of the data, focusing on grouping data based on its inherent characteristics.

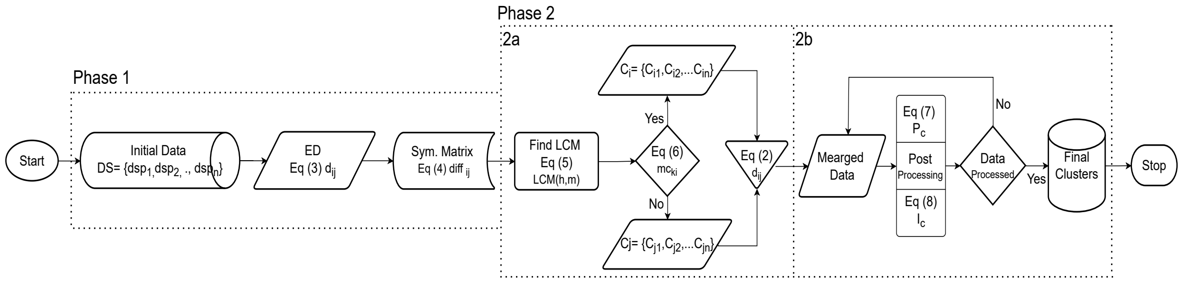


Figure 2: Flowchart: Phase 1 initializes data by computing the ED, resulting in a symmetric distance matrix. Phase 2 forms clusters by computing the LCM. Main clusters (C_i) and single-value clusters (C_j) are managed with ED, followed by post-processing to address repetitions (P_c) and shared characteristics (I_c), achieving the final clusters.

The proposed algorithm is illustrated in Figure 2 and consists of two phases. In Phase 1, the data initialization involves computing the Euclidean distance (ED) on the input data, creating a symmetric matrix of distances that serves as the primary input. In Phase 2, clusters are formed using the LCM calculated from the input data, and sample assignments are made accordingly. Here, C_i represents the main clusters, while C_j denotes the clusters containing independent values because of noise. The distance of C_j to the nearest clusters is defined to retain information and correctly place it in the nearest clusters. Post-processing addresses repetitive samples and sub-clusters, ensuring the accurate management of misplaced data points based on similar characteristics. The methodology follows the flowchart's order, which will be detailed in the next section.

3.1 Phase 1: Initial Processing of Input

3.1.1 Distance Scale Factor (DSF) ϑ

Initially, pairwise distances between samples are calculated and multiplied by the ϑ to refine clustering using Eq. (1). This hyperparameter is key to the LCM clustering technique, as it improves flexibility and adaptability by managing how clusters are formed. By adjusting ϑ, the algorithm fine-tunes its sensitivity to distances, influencing both the compactness and number of clusters. This also enables the technique to adapt to the specific characteristics of the samples, ensuring that meaningful cluster patterns are captured effectively. A visual depiction of ϑ's impact on clustering results is provided in Section 5.3.

$$d_{ij} = \sqrt{\sum_{k=1}^{n} (dsp_{ik} - dsp_{jk})^2 \times \vartheta} \tag{1}$$

The total number of estimated distances d_{ij} is determined by the number of distances between each point and all other points, which is comparable to the number of edges in a fully connected graph. This is calculated as $\frac{n(n-1)}{2}$, where n is the total number of points i.e., dsp. Results are stored in a symmetric matrix using Eq. (2) whose size is $n \times n$ dimension, which the proposed clustering algorithm will use as an input to generate clusters in the second phase.

$$diff_{ij} = \begin{bmatrix} 0 & d_{12} & d_{13} & \dots & d_{1n} \\ d_{12} & 0 & d_{23} & \dots & d_{2n} \\ d_{13} & d_{23} & 0 & \dots & d_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{1n} & d_{2n} & d_{3n} & \dots & 0 \end{bmatrix} \tag{2}$$

3.2 Phase 2: Proposed Clustering Technique

3.2.1 Phase 2a: Least Common Multiple (LCM)

The proposed algorithm is based on the calculation of the LCM. It is the smallest positive integer divisible by two or more integers. LCM of two numbers h and m can be calculated using Eq. (3).

$$LCM(h,m) = \frac{|h \cdot m|}{\gcd(h,m)} \tag{3}$$

Where gdc(h, m) indicates the greatest common divisor of h and m.

Algorithm 1 illustrates how the clustering technique operates. It begins by determining if the initial sample belongs to a cluster, which helps form the first cluster. Then, it examines the distance vectors of the samples from the $diff_{ij}$. At each step, the algorithm computes and compares the LCM of the existing clusters with the LCM of the new vector nv. The decision to assign a sample to a cluster is based on the membership condition mc_{ki} as shown in Eq. (4).

$$mc_{ki} = \begin{cases} 1, \text{ if } LCM_{nv+C_i} \quad \% \quad LCM_{C_i} == 0 \\ 1, \text{ if } LCM_{C_i} \quad \neq 0 \\ 1, \text{ if } LCM_{C_i} \quad \leq \quad \rho \\ 0, \text{ else} \end{cases} \tag{4}$$

When the condition in Eq. (4) is met, the algorithm adds a new sample vector nv to an existing cluster C_i. If not, a new cluster C_j is created for the single sample. This continues until all samples in $diff_{ij}$ are processed. The threshold ρ helps avoid dividing by zero and is typically set to 1 by default. Using these rules, the algorithm groups similar samples into similar clusters. Finally, C_j are combined with existing clusters C_i to prevent loss of information based on ED. Following this, the algorithm produced merged data for phase 2b.

3.2.2 Phase 2b1: Handling Repeating Values (P_c)

After merging the collected data, post-processing addresses increased sample sizes due to repeated sample points appearing across different clusters. Placement conditions P_c i.e., Eq. (5) ensure proper assignment of these repeated values. The first condition transfers a repeated value to the cluster where it appears more frequently. The second condition uses the nearest neighbours approach, prioritizing clusters with at least three nearby points to the repeated value, exceeding a threshold ς. For example, if a value appears more often in cluster A than in cluster B, it moves to A unless cluster B has more nearby neighbours. This ensures the effective allocation of repeated values, enhancing clustering accuracy and integrity.

$$P_c = \begin{cases} 1, \text{ if repeating value in C1 } > C2 \\ 1, \text{ if neighbours of repeating value in C1} \geq \varsigma \\ 0, \text{ else} \end{cases} \tag{5}$$

3.2.3 Phase 2b2: Handling Sub-Clusters (I_c)

Due to the influence of the indoor environment, certain MFS exhibit similar characteristics, forming distinctive sub-clusters (I_c) within primary clusters. These I_c share characteristics with other clusters, which is a primary research focus. To address this, we calculate the LCM of all clusters, excluding I_c, and divide it by the LCM of I_c, setting the ς range as $[0.5, 1.5]$ i.e., Eq. (6). This process iterates until final clusters are acquired, ensuring clustering is based on MFS characteristics rather than just distance metrics. This method automatically determines clusters' correct number and shape, distinguishing it from standard approaches like DBSCAN, agglomerative clustering, and K-means.

$$I_c = \begin{cases} 1, \text{ if } LCM_{I_c} \quad \% \quad LCM_{C_i} == 0 \\ 1, \text{ if } LCM_{I_c} \quad \% \quad LCM_{C_i} \in \varsigma \\ 0, \text{ else} \end{cases} \tag{6}$$

Algorithm 1: LCM Clustering Algorithm.

```
Input: n data points, clusters (initially
  empty)
Output: Clusters
for each data point n_i do
    if n_i is already assigned to a cluster then
        skip
    end
    for each existing cluster j do
        if Equation (9) is true (LCM of n_i and
          cluster j) then
            Assign n_i to cluster j break
        end
    end
    if i is not assigned to any cluster then
        create a new cluster with n_i as its only
          member
        for each unassigned data point k
          (starting from n_i + 1) do
            Calculate LCM of n_i and k
            if LCM is within threshold ρ then
                Add k to the new cluster
            end
        end
        Add the new cluster to clusters
    end
end
```

4 EVALUATION CRITERIA

The performance of the LCM clustering algorithm was evaluated using the state-of-the-art techniques described below.

4.1 State-of-the-Art Clustering Validity Index

We address three state-of-the-art evaluation criteria for the proposed algorithm: the Silhouette score (SS), Calinski-Harabasz Index ($CH-I$), and Davies-Bouldin Index ($DB-I$).

- Silhouette score: It is mathematically defined as

$$SS = \frac{1}{dsp} \sum_{i=1}^{dsp} \left(\frac{b_i - a_i}{\max(a_i, b_i)} \right) \tag{7}$$

  Where dsp is the total number of samples, a_i is the average distance between sample i and all other samples in the same cluster, and b_i is the average distance between sample i and all samples in the nearest neighbouring cluster.

- Calinski-Harabasz Index: It is mathematically defined as

$$CHI(K) = \frac{\left[\frac{\sum_{i=1}^{K} |C_i| d(v_i, v)^2}{K-1}\right]}{\frac{\sum_{i=1}^{K} \sum_{dsp \in C_j} d(dsp, v_i)^2}{dsp-K}} \tag{8}$$

Where V_i is the centroid of the C_i cluster and v is the global centroid of all the dsp in DS.

- Davies-Bouldin Index: This technique is mathematically defined as

$$DBI(K) = \frac{1}{K}\sum_{i=1}^{K}\max_{j\neq i}\left(\frac{\text{avg}(C_i)+\text{avg}(C_j)}{\xi(C_i, C_j)}\right) \tag{9}$$

where, in Eq. (9) $avg(C_i)$ is $\frac{1}{|C|}\sum_{dsp\in C_i} d(dsp, v_i)$, here v is the centroid of C cluster and $|C|$ are the number of samples in cluster C. $\xi(C_i, C_j) = d(v_i, v_j)$, where v_i is the centroid of cluster C_i.

4.2 Datasets for Evaluation

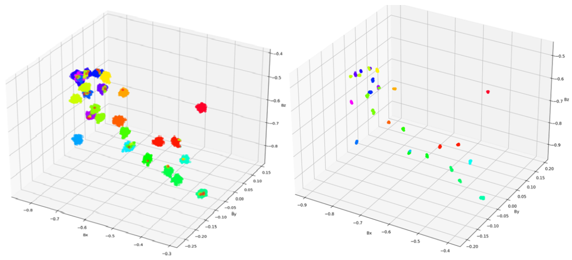
The proposed technique was evaluated using the benchmark datasets. One dataset, covering an area of 185 m^2 with 36 RPs, was collected at 10 Hz using a Sony Xperia M2 mobile phone (Barsocchi et al., 2016). Another dataset, obtained from a 60 m^2 area with 27 RPs, was gathered in an environment influenced by ferromagnetic materials, using Huawei P8 lite and iPhone 13 Pro Max devices as explained in (Rafique et al., 2023a). Data were collected at a frequency of 100 Hz and 10 Hz per second respectively, consisting of MFS components Bx, By, and Bz. These datasets allowed for comprehensive testing of the clustering technique on benchmark and real-time data, demonstrating its effectiveness in various conditions as shown in Table 1.

5 CLUSTER ASSESSMENT BASED ON STATE-OF-THE-ART EVALUATION INDICES

The clustering technique was evaluated on noisy (raw) and clean datasets using state-of-the-art evaluation indices as described in section 4.

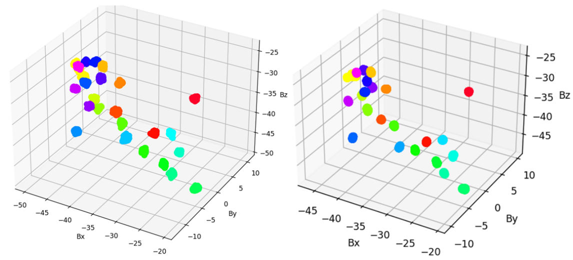
5.1 Sub-Clusters with Shared Characteristics I_c

This section addresses the research question and presents the issue of sample distribution with similar characteristics as discussed in Section 3.2.3 on a real



(a) I_c of Huawei (b) I_c of iPhone

Figure 3: Representation of I_c on both Data Sets.



(a) Huawei Dataset (b) iPhone Dataset

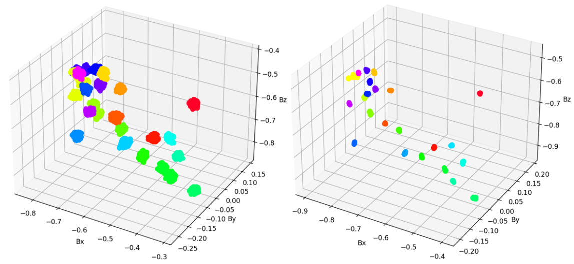
Figure 4: Clustering Patterns in Noisy Datasets.

dataset. I_c are clusters sharing common characteristics with other clusters. Samples with similar characteristics form sub-clusters I_c within a host cluster and are classified based on the highest shared characteristic values using Eq. (6). Figure 3 illustrates these sub-clusters within the Huawei and iPhone datasets, with each sub-cluster identifiable by colours matching their primary cluster. This indicates the presence of sub-clusters that share specific characteristics with other clusters while maintaining their unique identities.

5.2 Cluster Representation

5.2.1 Noisy Dataset

The raw data from each device represents noisy data, as shown in Figure 4. This data was fed into the LCM clustering approach to evaluate its handling of data



(a) Huawei Dataset (b) iPhone Dataset

Figure 5: Clustering Patterns in Clean Datasets.

noise. Figure 4a shows the clusters from the noisy Huawei dataset, and Figure 4b shows the clusters from the noisy iPhone dataset. The LCM approach successfully created balanced clusters with arbitrary shapes for both datasets.

5.2.2 Clean Data

The collected datasets underwent preprocessing to mitigate distortions, offsets, and noise in the magnetic field readings by following the strategy explained in (Rafique et al., 2023a), resulting in clean, noise-free datasets. These clean datasets were then used to evaluate the proposed clustering technique's efficacy compared to the noisy datasets. The resulting clusters from the processed datasets are shown in Figure 5, illustrating the positive impact of preprocessing on clustering outcomes.

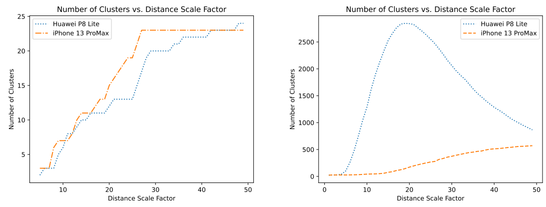
5.3 Fine Tuning the DSF ϑ

5.3.1 DSF and Noisy Dataset

Figure 6 illustrates the relationship between the number of clusters and ϑ. For clean datasets, the algorithm generated 21 and 23 clusters and noisy datasets exhibited unusual behaviour, with the number of clusters initially constant at 23, increasing as ϑ increased. Experiments identified ϑ values of 1 or 2 as optimal for clustering noisy datasets, producing favourable clusters as shown in Figure 7. Despite the increasing ϑ, the number of clusters remained constant at 23 with optimal evaluation results as in Figure 8. This indicates the algorithm's ability to generate precise clusters with favourable evaluation scores at these optimal ϑ values. Figures 8 provide a comprehensive analysis of the algorithm's performance for different ϑ values and the number of clusters generated.

5.3.2 DSF and Clean Dataset

After numerous experiments, the optimal values of ϑ for clean datasets were found to be in the range of 55 to 100. The selection of ϑ is crucial for adjusting the data magnitude and achieving effective clustering,



(a) Both Clean Datasets (b) Both Noisy Datasets

Figure 6: Exploration of number of clusters Vs ϑ for clean and noisy datasets.

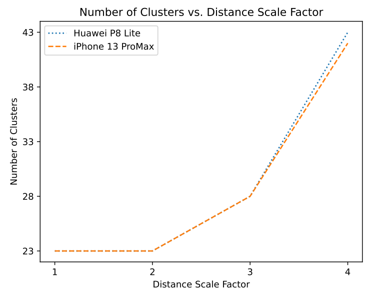
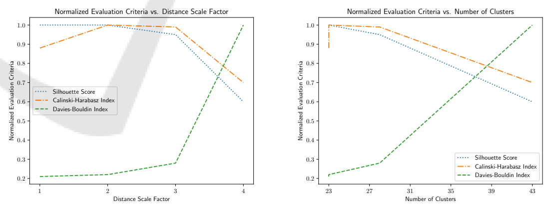


Figure 7: Close examination of ϑ Vs. number of clusters for the noisy data as compared to Figure 6b.

considering variations in magnitude during data calibration. Figure 9 shows the relationship between ϑ and the evaluation criteria for clean datasets. It illustrates how the evaluation metrics change as ϑ increases, highlighting its effect on clustering performance. Additionally, Figure 10 demonstrates the unusual clustering behaviour on noisy datasets. Clustering at ϑ of 2 yields favourable evaluation scores.

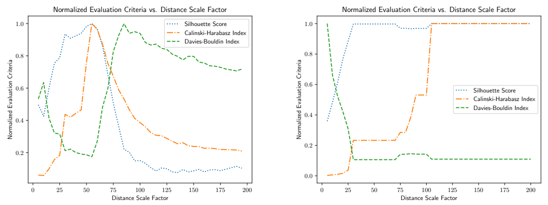


(a) (b)

Figure 8: (a) It presents the optimal value of DSF for clustering noisy data based on evaluation techniques. (b) Presents the number of clusters i.e., 23 based on the optimal value of DSF using evaluation metrics.

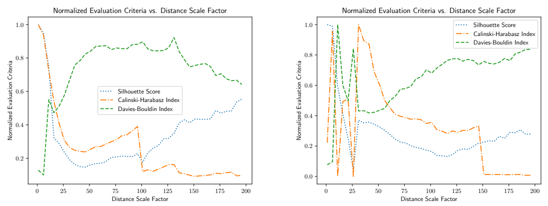
Table 1: Evaluation Matrix across Various Datasets, Devices, and Study Environments Using Three Evaluation Metrics. An SS close to one indicates a strong cluster, a higher CH-I reflects better-defined clusters and a DB-I closer to zero signifies strong cluster separation.

SR. No.	Data set	Total Sample	SS [-1,1]	CH-I (high)	DB-I [-0,1]
1	Huawei P8 (Noisy)	8920	0.83	229098.30	0.21
2	iPhone 13 Pro Max (Noisy)	1882	0.91	229154.46	0.11
3	Huawei P8 (Clean)	8920	0.72	60817.75	0.30
4	iPhone 13 Pro Max (Clean)	1882	0.91	94257.08	0.11
5	Sony Xperia M2	36795	0.99	72902.23	0.21



(a) Clean Huawei Dataset (b) Clean iPhone Dataset

Figure 9: (a) It presents the optimal value of DSF for clustering clean data based on evaluation techniques. (b) Presents the number of clusters obtained by the optimal value of DSF using evaluation metrics.



(a) Noisy Huawei Dataset (b) Noisy iPhone Dataset

Figure 10: Exploration of evaluation techniques on noisy data set vs DSF.

Table 1 displays the results of state-of-the-art evaluation techniques applied to various benchmark datasets. The obtained results fall within the defined threshold values of the techniques. A silhouette score of 0.5 or higher indicates strong clustering, with an ideal score being 1, while a score below 0 indicates weak clustering. The Calinski-Harabasz Index measures the ratio between-cluster dispersion to within-cluster dispersion, with a larger ratio indicating more well-defined clusters. The Davies-Bouldin Index compares within-cluster distances to between-cluster distances and is bounded between -0 and 1, with a lower score being preferable.

The computational complexity of the proposed algorithm comprises two main parts: calculating the pairwise Euclidean distance and the clustering process. The Euclidean distance calculation has a time complexity of $O(n^2)$ due to a nested loop over 'n' data points. The clustering algorithm, which includes comparisons, loops, and least common multiple calculations, has a combined time complexity of $O(n^2+n)$. In the worst-case scenario, where each data point must be compared with all existing clusters and potentially form a new cluster, the algorithm performs the maximum number of comparisons and assignments.

6 CONCLUSION

This study presents a novel clustering approach, characteristic-based least common multiple (LCM) clustering, that aims to improve indoor localization accuracy. This method effectively identifies clusters with varied densities, shapes, and sizes by leveraging sample similarity and magnetic field sensor properties.

The LCM-based clustering process starts with calculating pairwise distances and constructing a symmetric matrix. Clusters are then formed by calculating the LCM of sample attributes, allowing new points to join existing clusters or form new ones based on LCM criteria. The algorithm also merges independent clusters into neighbouring ones based on minimum distance requirements.

A key feature of this approach is its ability to detect misplaced sub-clusters within larger clusters and reassign them to the correct clusters. This improves the identification of distinct entities and reduces prediction ambiguity, leading to a significant boost in positioning accuracy.

The algorithm was tested on both noisy and clean datasets, as well as benchmark datasets. The proposed method demonstrated strong clustering performance, as confirmed by evaluation metrics such as the Silhouette Score, Calinski-Harabasz Index, and Davies-Bouldin Index.

In the future, we plan to apply this technique to real-world datasets in diverse, complex environments to assess its effectiveness in practical indoor localization scenarios.

ACKNOWLEDGEMENT

This work is financially supported from: PNRR MUR project PE0000013-FAIR (Future Artificial Intelligence Research) and MUR project ARS01_00592 reCITY - Resilient City Everyday Revolution.

REFERENCES

Ankerst, M., Breunig, M. M., Kriegel, H.-P., and Sander, J. (1999). Optics: Ordering points to identify the clustering structure. *ACM Sigmod record*, 28(2):49–60.

Anuwatkun, A., Sangthong, J., and Sang-Ngern, S. (2019). A diff-based indoor positioning system using fingerprinting technique and k-means clustering algorithm. In *2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pages 148–151. IEEE.

Barsocchi, P., Crivello, A., La Rosa, D., and Palumbo, F. (2016). A multisource and multivariate dataset for indoor localization methods based on wlan and geomagnetic field fingerprinting. In *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–8. IEEE.

Berahmand, K., Mohammadi, M., Faroughi, A., and Mohammadiani, R. P. (2022). A novel method of spectral clustering in attributed networks by constructing parameter-free affinity matrix. *Cluster Computing*, pages 1–20.

Cui, Z., Jing, X., Zhao, P., Zhang, W., and Chen, J. (2021). A new subspace clustering strategy for ai-based data analysis in iot system. *IEEE Internet of Things Journal*, 8(16):12540–12549.

El Khediri, S., Fakhet, W., Moulahi, T., Khan, R., Thaljaoui, A., and Kachouri, A. (2020). Improved node localization using k-means clustering for wireless sensor networks. *Computer Science Review*, 37:100284.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pages 226–231.

Fang, S.-G., Huang, D., Cai, X.-S., Wang, C.-D., He, C., and Tang, Y. (2023). Efficient multi-view clustering via unified and discrete bipartite graph learning. *IEEE Transactions on Neural Networks and Learning Systems*.

Junyi, G., Li, S., Xiongxiong, H., and Jiajia, C. (2021). A novel clustering algorithm by adaptively merging subclusters based on the normal-neighbor and merging force. *Pattern Analysis and Applications*, 24(3):1231–1248.

Lee, S. G. and Lee, C. (2020). Developing an improved fingerprint positioning radio map using the k-means clustering algorithm. In *2020 International Conference on Information Networking (ICOIN)*, pages 761–765. IEEE.

Mavridis, L., Nath, N., and Mitchell, J. B. (2013). Pfclust: a novel parameter free clustering algorithm. *BMC bioinformatics*, 14:1–21.

Paparrizos, J. and Gravano, L. (2015). k-shape: Efficient and accurate clustering of time series. In *Proceedings of the 2015 ACM SIGMOD international conference on management of data*, pages 1855–1870.

Rafique, H., Almagrabi, A. O., Shamim, A., Anwar, F., and Bashir, A. K. (2020). Investigating the acceptance of mobile library applications with an extended technology acceptance model (tam). *Computers & Education*, 145:103732.

Rafique, H., Patti, D., Palesi, M., and catania, V. (2023a). m-bmc: Exploration of magnetic field measurements for indoor positioning using mini-batch magnetometer calibration. In *2023 First IEEE International Conference on Mobility: Operations, Services, and Technologies (MOST)*, pages 55–61. IEEE.

Rafique, H., Patti, D., Palesi, M., La Delfa, G. C., and Catania, V. (2023b). Optimization technique for indoor localization: A multi-objective approach to sampling time and error rate trade-off. In *2023 IEEE Third International Conference on Signal, Control and Communication (SCC)*, pages 01–06. IEEE.

Rafique, H., Ul Islam, Z., and Shamim, A. (2023c). Acceptance of e-learning technology by government school teachers: Application of extended technology acceptance model. *Interactive Learning Environments*, pages 1–19.

Ren, J., Wang, Y., Niu, C., Song, W., and Huang, S. (2019). A novel clustering algorithm for wi-fi indoor positioning. *IEEE Access*, 7:122428–122434.

Singh, M. and Soni, S. K. (2019). Fuzzy based novel clustering technique by exploiting spatial correlation in wireless sensor network. *Journal of Ambient Intelligence and Humanized Computing*, 10:1361–1378.

Vinciguerra, E., Russo, E., Palesi, M., Ascia, G., and Rafique, H. (2024). Improving lstm-based indoor positioning via simulation-augmented geomagnetic field dataset. In *2024 IEEE International Conference on Mobility, Operations, Services and Technologies (MOST)*, pages 251–259. IEEE.

Vo-Van, T., Nguyen-Hai, A., Tat-Hong, M., and Nguyen-Trang, T. (2020). A new clustering algorithm and its application in assessing the quality of underground water. *Scientific Programming*, 2020:1–12.

Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and computing*, 17:395–416.

Xu, Y., Huang, D., Wang, C.-D., and Lai, J.-H. (2024). Deep image clustering with contrastive learning and multi-scale graph convolutional networks. *Pattern Recognition*, 146:110065.