I.J. Information Engineering and Electronic Business, **2024, 5, 95-109**
Published Online on October 8, 2024 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijieeb.2024.05.06

Dynamic Data Aggregation Model for Social Internet of Things Devices: Exploring the Static and Mobile Nature

Meghana J.
Department of Studies in Computer Science, University of Mysore, Mysuru-570006, Karnataka, India
E-mail: meghanajgowda1984@gmail.com
ORCID iD: https://orcid.org/0009-0004-4855-5993

Hanumanthappa J.
Department of Studies in Computer Science, University of Mysore, Mysuru-570006, Karnataka, India
E-mail: hanumsj@yahoo.com
ORCID iD: https://orcid.org/0000-0002-6031-6993

S. P. Shiva Prakash*
Department of Information Science and Engineering, JSS Science and Technology University, Mysuru-570006, Karnataka, India
E-mail: shivasp@jssstuniv.in
ORCID iD: https://orcid.org/0000-0003-3490-6292
*Corresponding Author

Kirill Krinkin
School of Computer Science and Engineering, Constructor University, Bremen, Germany
E-mail: kirill@krinkin.com
ORCID iD: https://orcid.org/0000-0001-5949-7830

Received: 13 January, 2024; Revised: 23 February, 2024; Accepted: 12 March, 2024; Published: 08 October, 2024

Abstract: The increasing ubiquity of Social Internet of Things (SIoT) devices necessitates innovative data aggregation techniques to distill meaningful insights from diverse sources. This study introduces a Dynamic Data Aggregation Model for SIoT devices. The model aims to amalgamate static and mobile device data, employing Density-Based Spatial Clustering of Applications with Noise (DBSCAN) for spatial clustering and Recurrent Neural Networks (RNN) for predicting mobile device movement patterns. The purpose is to offer a holistic approach to predictive analytics in the SIoT domain by seamlessly integrating these methodologies. The model begins with data preprocessing, ensuring data quality. It then applies DBSCAN for spatial clustering, enabling a comprehensive understanding of spatial relationships between devices. Simultaneously, RNNs are used for predictive modeling, specifically in forecasting mobile device movement patterns. The integration of DBSCAN clustering and RNNs forms the model's innovative core, providing a unified solution for dynamic data aggregation. Comprehensive testing demonstrates the model's notable accuracy in predicting mobile device movement patterns. By combining the spatial clustering capabilities of DBSCAN with the predictive power of RNNs, the model effectively unifies static and mobile data, advancing predictive analytics in the SIoT context. The proposed model yielded average values of 0.14604 (Mean Squared Error), 2.678385 (Mean Absolute Error), 0.307154 (Root Mean Squared Error), and 1.342317 (Mean Absolute Percentage Error), respectively. The Dynamic Data Aggregation Model proves its efficacy in addressing SIoT challenges. The integration of DBSCAN clustering and RNNs offers a versatile framework for dynamic data analysis, contributing significantly to predictive analytics in SIoT contexts.

Index Terms: DBSCAN, RNN, Data Aggregation, Social Internet of Things

1. Introduction

The Social Internet of Things (SIoT) presents a landscape where interconnected devices engage in seamless communication and collaboration, aiming to enhance various aspects of daily life. In this dynamic network, effective data aggregation stands as a crucial factor in unlocking the full potential of the interconnected devices. SIoT environments are marked by a diverse array of devices continuously generating dynamic data streams.

The primary challenge within SIoT lies in aggregating, processing, and extracting meaningful insights from this heterogeneous data. This data encompasses not only traditional Internet of Things (IoT) metrics but also socially driven interactions. Social interactions within SIoT ecosystems introduce an additional layer of complexity to the data analysis process. Conventional models face difficulties in contextualizing and interpreting the subtleties inherent in human-device and device-device social relationships. This limitation hampers the development of intelligent systems capable of adapting to social contexts. Within the SIoT landscape, various relationship types contribute to the complexity of interactions such as Parent-Object Relationships, Social-Object Relationships, Co-work Object Relationships, Co-location Object Relationships, Strange-Object Relationships, and more. Each relationship type signifies a different aspect of interaction, adding richness to the social dynamics within the net- work. As SIoT networks grow in scale, the need for scalable solutions that can handle the increasing volume of data while maintaining real-time responsiveness becomes paramount. Existing frameworks face challenges in efficiently scaling with the expanding network and ensuring timely data processing. Currently, there is a gap in the availability of comprehensive dynamic data aggregation models specifically tailored for SIoT environments. Existing solutions may lack the versatility required to capture the full spectrum of social interactions and diverse IoT data sources. This paper introduces a novel Dynamic Data Aggregation Model tailored specifically for SIoT environments, addressing the intricacies associated with the fusion of social interactions and the Internet of Things. In the realm of SIoT, devices are not mere inanimate objects; they are active participants in social ecosystems, contributing to and deriving insights from the collective intelligence of the network. As such, the conventional models for static and mobile devices may fall short in capturing the nuances of social dynamics intertwined with IoT functionalities. Our proposed model integrates sophisticated algorithms, drawing inspiration from social network analysis and IoT data processing, to provide a holistic solution for dynamic data aggregation in this unique context. The amalgamation of interconnected devices and social interactions presents a unique set of challenges regarding data aggregation and analysis. Traditional data processing models often fall short of capturing the intricacies of dynamic social relationships within the IoT framework.

1.1 Key contributions of the study:

- The study introduces an innovative Dynamic Data Aggregation Model tailored for SIoT devices, seamlessly integrating static and mobile data.
- The model's unique combination of Density-Based Spatial Clustering (DBSCAN) and Recurrent Neural Networks(RNNs) enhances predictive analytics in SIoT by offering a cohesive framework for dynamic data aggregation.
- Provided a versatile tool for accurate prediction of mobile device movements and fostering a clearer understandingof spatial relationships between devices.

The rest of the paper is structured as follows: Section 2 reviews related works conducted by different authors. Section 3 defines the problem statement, while Section 4 introduces the proposed method, including the system model, methodology design, and algorithm. The findings are elaborated in Section 5, and the paper concludes with Section 6, which includes the conclusion and outlines future work.

2. Related Works

The ensuing review offers an outline of pertinent studies within the realm of dynamic data aggregation, predictive analytics, and the amalgamation of spatial clustering and machine learning within the framework of the SIoT. Quentin Bramas and colleagues [1] focused on addressing the complexity of aggregating both static and dynamic data, aiming to minimize the data aggregation time. Al-kahtani et al. [2] successfully introduced a multilayer technique for aggregating data in big data frameworks, particularly those involving sensor-based devices. This technique not only reduces overall energy consumption in the network but also minimizes latency in data transmission. In their work, J. W. Raymond et al. [3] explored Cooperative Communication in Machine-to-Machine (M2M) communication, dependent on protocols such as clustering and communication. The challenges discussed include the complexity of design configuration, in-creased message overheads, network expansion (inter and intra), and issues related to delay and data repetitiveness based on communication and clustering protocols. They also proposed enhancements like full-duplex communication, non-regenerative hand-off methods, cross-layer improvements, and a clustering scheme. Tabinda Salman et al. [4] provided detailed insights into upcoming trends in designing cellular networks for IoT devices, introducing the concept of Machine Type Communications (MTC). Their analysis considered the application's goals and challenges, proposing solutions for efficient data access and various data aggregation methods. Chang-Sik Choi et al. [5] proposed a system to understand the architecture of wireless devices, specifically those collecting data from vehicles. Their examination focused on the development of the coverage area over time, evaluating minimal delay and time required for identifying devices on the roadside. Ogud et al. [6] explored the global concept of IoT communication in machine-type, conducting experiments on mobile devices' performance within cellular networks. The results provided insights into managing traffic flow in the network.

Juan Manuel Rodriguez et al. [7] aimed to identify a cell phone's battery charge status based on its connection to AC, WIFI, and screen usage. They utilized Recurrent Neural Network algorithms for predicting the phone's upcoming state, considering the past states. Tae Won Ban et al. [8] addressed challenges in Device-to-Device (D2D) networks, proposing a novel transmission algorithm based on Deep Learning techniques with Convolutional Neural Networks (CNN). Their experiments demonstrated the algorithm's effectiveness in managing radio resources. Tianfu Wang et al. [9] introduced the Network Space DBSCAN (NS-DBSCAN) algorithm, providing a new method for clustering based on density and specifying cluster structures. Evaluation results indicated superior performance compared to other clustering algorithms. Yuelei Xiao et al. [10] proposed a vehicle location prediction algorithm, combining Long Short-Term Memory (LSTM) and spatiotemporal feature transformation. The algorithm successfully reduced data loss and improved location identification accuracy compared to previous models. Nama et al. [11] presented a three-way solution for traffic congestion reduction, involving data gathering techniques, accuracy in obtained data through machine learning algorithms, and information on different traffic scheduling strategies. Kim T et al. [12] focused on predicting human motion signals obtained from attached sensors for medical service purposes. Their proposed method, combining plot techniques and neural networks, outperformed conventional methods in identifying accidents faster. Choi et al. [13] proposed the Multidimensional Spatiotemporal data – Density-Based Spatial Clustering Applications with Noise (MDST-DBSCAN) method for efficient clustering of large, multidimensional data. The model demonstrated accurate cluster identification within the given time. Mazin Hameed et al. [14] introduced the Distributed Density-Based Spatial Clustering of Applications with Noise (DDSCAN) protocol for energy-saving in IoT sensor devices. Their systematic strategy identified cluster heads based on energy, nearby residents, and interval information, proving more efficient than previous methodologies. Yasser Nabil et al. [15] developed a mathematical model to analyze the interaction between data granularity, transmission delay, and reliability. The model aimed to optimize packet size, power supply, antenna directivity, and transmission rate,enhancing overall device performance.

Kang Tan et al. [16] successfully controlled a vehicular network, addressing handover management, resource efficiency, and decision-making related to network formation/deformation. They also discussed the application of machine learning in vehicular networks. Lucy Dash et al. [17] proposed the spatial and Temporal Correlation-based Data Redun dancy Reduction (STCDRR) protocol, demonstrating a 7.2 times improvement in data redundancy removal compared to other algorithms. The protocol achieved higher data compression rates and reduced false data. Khattak et al. [18] developed an IoT-enabled optimal data aggregation method for urban surveillance, converting raw information into refined data with minimal information loss. The proposed approach outperformed previous methods in terms of refinement pro- portion, information loss, energy efficiency, and lifespan. Seo Jin Chang et al. [19] introduced the Delay-Based Dynamic Clustering (DBDC) method, relying on DBSCAN to reduce communication delay in large control clusters. Evaluation results showed the method's success in planning clusters for optimal control execution. Sakorn Mekruksavanich et al. [20] combined a model for human activity recognition with real-time dataset testing, showcasing improved identification of various human activities using gyroscopes and accelerometers in a digital watch. Dae Young Kim et al. [21] proposed an aggregation data model for identifying the proper node of aggregation and describing data transmission methods. Their evaluation demonstrated the model's superior performance in simulated data scenarios. Erskine SK et al. [22] presented the Secure Data Aggregation Authentication (SDAA) protocol, minimizing delay and increasing energy efficiency in underwater vehicular wireless sensor networks. The method proved effective in improving energy efficiency and reducing network delays compared to previous approaches. Xiaoya An et al. [23] introduced the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Spatial-Temporal Random Partitioning (STRP-DBSCAN) algorithms for clustering spatial-temporal data. The STRP-DBSCAN algorithm significantly reduced clustering time compared to DBSCAN, and an additional Prioritizes Experience Replay-Soft Actor-Critic (PER-SAC) algorithm further improved clustering accuracy and stability.

The current literature addresses a broad spectrum of subjects encompassing data aggregation, communication proto- cols, machine learning applications, and clustering algorithms, demonstrating notable advancements and inventive solutions across diverse IoT-related domains. However, it is noteworthy that a gap exists in the exploration of data aggregation for Static and IoT (SIoT) devices, taking into account both their static and mobile nature and establishing inter-device relationships.

3. Problem Statement

In the realm of SIoT, conventional data processing falls short of capturing the dynamic and socially driven nature of device interactions. The challenge lies in aggregating heterogeneous data streams, understanding nuanced social dynamics, and ensuring scalability concerns. Existing models lack the specificity required for SIoT. This prompts the need for an innovative dynamic data aggregation approach tailored to the intricacies of SIoT, ensuring efficient processing, contextual understanding of social interactions, and scalability in a landscape where interconnected devices and social elements converge.

4. Proposed method

4.1 System Model

Consider a system with a set of static devices, denoted as $D = \{D_1, D_2, \ldots, D_N\}$, and mobile devices represented as $M = \{M_1, M_2, \ldots, M_M\}$. Let X_t be the dynamic dataset generated by all devices at time t, where $X_t^{(i)}$ is the data generated by device D_i and $X_t^{(j)}$ is the data generated by mobile device M_j. The spatial locations of devices are defined as $C_i = (x_i, y_i)$ for static devices and $C_j = (x_j, y_j)$ for mobile devices. Let R be the set of social relationships$r_{ij} \mid d_i, d_j \in D$.

4.1.1 Objective Function

The objective is to minimize the data aggregation cost while considering spatial clustering and predictive analytics.

The cost function is defined as follows:

$$Cost = \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} .\ d\left(C_i, C_j\right) .\ X_t^{(i)} .\ X_t^{(j)} \tag{1}$$

Where w_{ij} is a weight factor representing the influence between the static device, D_i and mobile device M_j and $d(C_i, C_j)$ Is the distance function between the spatial coordinates of devices, D_i, M_j, $X_t^{(i)}$ and $X_t^{(j)}$ are the data generated byDevices D_i and M_j at time t.

4.1.2 Proposed Solution

The proposed solution involves employing Density-Based Spatial Clustering of Applications with Noise (DBSCAN) for spatial clustering and Recurrent Neural Networks (RNN) for predicting mobile device movement patterns. The clusters are formed by identifying the devices having the same relationship that are obtained through certain constraints like device protocols, device ID, device models etc.

Social Relationship Constraints

$$\sum_{j \in D} r_{ij} = 1 \ \forall_i \tag{2}$$

$$\sum_{i \in D} r_{ij} = 1 \ \forall_j \tag{3}$$

DBSCAN for Spatial Clustering: DBSCAN is utilized to cluster devices based on their spatial proximity.

Let $C_1, C_2, \ldots, C_{NC}$ represent the clusters formed by DBSCAN. The inclusion of DBSCAN is beneficial for identifying groups of devices that are spatially close, providing context for potential social interactions. Mathematically, DBSCAN assigns each device d_i to a cluster C_j based on spatial coordinates (x_i, y_i). This can be expressed as:

$$C_i = \{d_j \mid distance((x_i, y_i), (x_j, y_j)) \le \epsilon\} \tag{4}$$

Where, ϵ is the radius within which devices are considered neighbors.

RNN for Movement Prediction: RNNs are employed to predict device movements over time intervals. The RNN model is trained on historical data, considering parameters such as start timestamp, stop timestamp, user ID, and spatial coordinates. Mathematically, the RNN model can be represented as a set of equations capturing the temporal dependencies in device movements:

$$h_t = RNN(h_{t-1}, input_t) \tag{5}$$

Where, h_t is the hidden state at time t, and $input_t$ is the input vector containing information about the device, timestamp,and spatial coordinates at time t. The predicted movement's $\hat{x}, \hat{y}$ can be obtained from the output of the RNN.

Utility Function Incorporating Clustering and Prediction: The utility function $f(d_i, d_j, t_k, p_l, r_{ij})$ is enhanced to consider both clustering information from DBSCAN and move-ment predictions from RNN.

Let $U_{ijk,l}$ denote the enhanced utility:

$$U_{ijk,l} = \alpha \cdot f(d_i,\ d_j,\ t_k,\ p_l,\ r_{ij}) + \beta \cdot DBSCAN\ Similarity(d_i, d_j) + \gamma \cdot RNN\ Prediction\ Accuracy(d_i,\ t_k,\ p_l) \quad (6)$$

Where, α, β, γ are weights determining the importance of each component. Let S_{ij} be a binary decision variable representing whether data from static device D_i is aggregated with data from mobile device M_j. The mathematical model for the proposed solution is formulated as an optimization problem:

$$\min_{S_{ij}} \cos t \quad (7)$$

subject to

$$\sum_{j=1}^{M} S_{ij} = 1 \quad \forall_i, \quad (8)$$

$$\sum_{j=1}^{N} S_{ij} = 1 \quad \forall_j, \quad (9)$$

$$S_{ij} \in \{0,1\} \quad \forall_{i,j} \quad (10)$$

Equation (2) ensures that each static device's data is aggregated with exactly one mobile device, and Equation (3) ensures that each mobile device receives data from at most one static device. The decision variables S_{ij} are binary, indicating whether the corresponding data aggregation is scheduled.

4.2 Methodology

The purpose of the Dynamic Data Aggregation Model is to provide a comprehensive solution for predictive analytics in the realm of the SIoT. Figure 1 shows the methodology proposed to aggregate the data generated by the SIoT devices. The data sources include static devices denoted as Device A, Device B, and Device C, alongside mobile counterparts identified as Device X, Device Y, and Device Z. The Data Aggregation Engine operates on two pivotal methodologies: Density-Based Spatial Clustering of Applications with Noise (DBSCAN) for robust spatial clustering, and Recurrent Neural Networks (RNNs) tasked with predicting mobile device movement patterns through a specific architecture and training configuration. The Communication Module integrates diverse protocols such as Bluetooth, Zig bee, Wi-Fi, and GSM, fostering effective data exchange and relationship establishment between devices. For storage, a dedicated repository is employed to store aggregated data in a structured manner for future analysis. Specific relationship types, including Parent-Object, Social-Object, Co-work Object, Co-location Object, and Strange-Object, are defined to characterize associations between devices within the SIoT ecosystem. In essence, this model offers a holistic solution that integrates spatial clustering, neural network predictions, and relationship characterization for a robust SIoT data aggregation system.

The model amalgamates data from both static and mobile devices, fostering a holistic approach to understanding device dynamics in diverse scenarios. The model employs two key methodologies:

4.2.1 Density-Based Spatial Clustering of Applications with Noise (DBSCAN):

This method is utilized for robust spatial clustering of device data. The parameters include epsilon (ε) set to 10 and a minimum number of samples required for a cluster (min samples) set to 3. The result is a meaningful clustering of devices based on their spatial proximity.

4.2.2 Recurrent Neural Networks (RNNs):

RNNs are employed to predict mobile device movement patterns. The model's architecture includes a SimpleRNN layer with 10 units and a dense output layer with 1 unit. The relu activation function is used. The model is trained for 100 epochs with the Adam optimizer and mean squared error (MSE) as the loss function.

4.3 Unique Aspects of Integrating DBSCAN and RNN within SIoT

4.3.1 Comprehensive Data Handling:

- DBSCAN addresses the spatial clustering of static devices, allowing the model to understand the spatial relationships and groupings among them.
- RNNs complement this by capturing the temporal dynamics of mobile devices, predicting their movement patterns over time.
- By integrating both methodologies, the model can effectively handle the diverse data characteristics present in SIoT environments, covering both static and mobile aspects.

4.3.2 Spatial-Temporal Fusion:

- The integration of DBSCAN and RNN enables a fusion of spatial and temporal information.
- DBSCAN captures spatial proximity, identifying clusters of devices based on their physical locations.
- RNNs capture temporal patterns, predicting how mobile devices move and interact with each other over time.
- This fusion allows for a more holistic understanding of device dynamics in SIoT scenarios, where both spatial and temporal aspects are crucial.

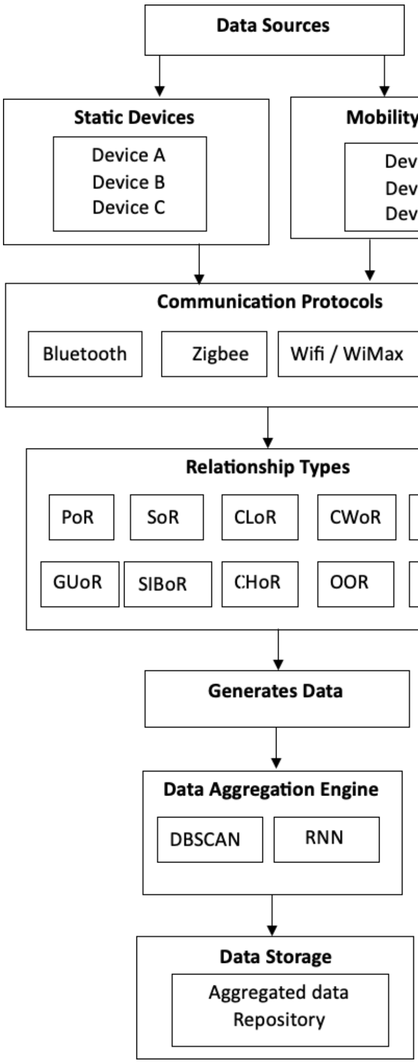

Fig. 1. Proposed methodology design

4.3.3 Robust Clustering and Prediction:

- DBSCAN's robust clustering algorithm ensures that static devices are effectively grouped based on spatial proximity, even in the presence of noise or outliers.
- RNNs, with their ability to capture sequential patterns, provide accurate predictions of mobile device movements, considering past trajectories and behaviors.
- This combination results in a robust model that can handle the inherent variability and unpredictability of SIoT data, enhancing the reliability of both clustering and prediction tasks.

4.3.4 Adaptability to Dynamic Environments:

- SIoT environments are dynamic, with devices constantly moving and interacting with each other.
- The integration of DBSCAN and RNNs allows the model to adapt to these dynamic changes by continuously updating spatial clusters and refining movement predictions based on the latest data.
- This adaptability ensures that the model remains effective in diverse SIoT scenarios, where device mobility and interactions play a crucial role.

Thus, the integration of DBSCAN and RNN within the SIoT context offers a unique approach that combines spatial clustering with temporal prediction, enabling a comprehensive understanding of device dynamics.

This integration facilitates robust data handling, spatial-temporal fusion, and adaptability to dynamic environments, making it well-suited forvarious applications in the realm of the Social Internet of Things.

4.4 Implications on the Broader Field of SIoT

The integration of DBSCAN and RNN methodologies within the SIoT context presents significant implications for the broader field:

- **Enhanced Understanding of Device Dynamics:** By combining spatial clustering and temporal prediction, the model facilitates a deeper understanding of device behaviors in SIoT environments, enabling more informed decision- making processes.
- **Improved Data Handling in Dynamic Environments:** The adaptability of the integrated model to dynamic changes ensures robust data handling, which is crucial in real-world SIoT scenarios where device mobility and interactions constantly evolve.
- **Potential for Advanced Applications:** The fusion of spatial-temporal information opens avenues for advanced SIoT applications such as smart city management, intelligent transportation systems, and environmental monitoring, where comprehensive data analysis is essential.

Thus, the integration of DBSCAN and RNN within the SIoT context offers a unique approach that combines spatial clustering with temporal prediction, enabling a comprehensive understanding of device dynamics. This integration facilitates robust data handling, spatial-temporal fusion, and adaptability to dynamic environments, making it well-suited for various applications in the realm of the Social Internet of Things. It also promises significant advancements in understanding device dynamics and enabling innovative applications across various domains.

4.5 Algorithm

The algorithm 1 processes a set of devices denoted as C, a data matrix at time t represented by X_t, and a weight matrix W. It incorporates the DBSCAN clustering algorithm with user-defined parameters, ε for distance and $min_samples$ for the minimum number of samples required for clustering. Additionally, an RNN (Recurrent Neural Network) is utilized in the algorithm. The primary objective is to dynamically aggregate data from clusters formed by DBSCAN, updating weights based on RNN predictions. The algorithm begins by applying DBSCAN to the set of devices C, resulting in clusters. Subsequently, it initializes a matrix denoted as $Aggregated_X$ with zeros. It then iterates over each cluster, calculating the centroid of devices within the cluster. Data matrix X_{cluster} and weight matrix W_{cluster} are extracted based on the devices in the cluster. The weights are then updated using RNN predictions for the corresponding data. Finally, the algorithm aggregates the results into $Aggregated_X$ using a specified formula. The output of the algorithm is the dynamically aggregated data matrix $Aggregated_X$, which reflects the contributions of different clusters based on their centroids and the associated RNN-informed weight updates.

Algorithm 1: Dynamic Data Aggregation Model with DBSCAN and RNN

Data: C: Set of devices, X_t: Data matrix at time t, W : Weight matrix, ε: DBSCAN parameter, min samples: Minimum samples for DBSCAN, RNN: Recurrent Neural Network
Result: Aggregated X: Aggregated data matrix
1 Apply DBSCAN to C with parameters ε and minsamples, obtaining clusters;
2 Initialize Aggregated Xmatrix with zeros;
3 **for** *each cluster* $C_{cluster}$ *in clusters* **do**
4 Calculate the centroid C_{centroid} of devices in C_{cluster};
5 Extract data matrix X_{cluster} and weight matrix W_{cluster} for devices in C_{cluster};
6 Update W_{cluster} using RNN predictions for X_{cluster};
7 Aggregate the result into AggregatedXusing the formula: AggregatedX += $W_{\text{cluster}} \cdot X^T \cdot X_t$;
8 **return** *Aggregated X*

5. Result and Discussion

5.1 Experiment

The model's effectiveness is evaluated through simulation using the Small World in Motion (SWIM) mobility model. Parameters include perceptual radius, user count, and motion values, providing realistic scenarios for static and mobile device interactions. The data generation process involves creating a dataset of 16 devices having 8 private and 8 public devices thatare in both static and mobility in nature. The 2D and 5D coordinates are randomly generated.

The Dynamic Data Aggregation Model showcases its effectiveness in providing a nuanced understanding of device dynamics, seamlessly unifying static and mobile data. This research contributes to advancing predictive analytics in SIoT contexts, offering

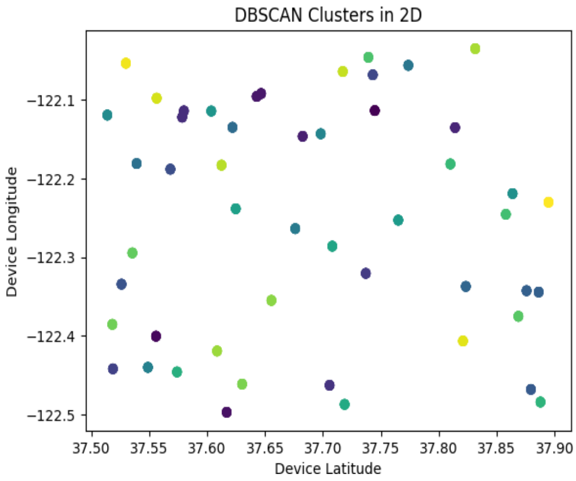

Fig. 2. DBSCAN clusters in 2D

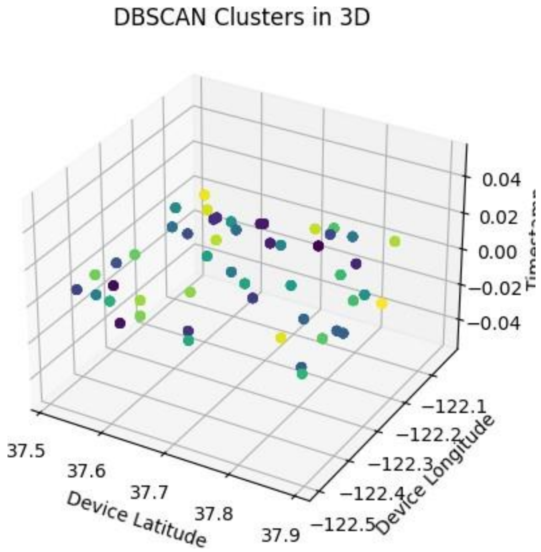

Fig. 3. DBSCAN clusters in 3D

A versatile framework for dynamic data analysis. The model demonstrates noteworthy accuracy in predicting mobile device movement patterns through comprehensive testing. The fusion of DBSCAN clustering and RNNs serves as the cornerstone of the model's novelty, offering a unified solution for dynamic data aggregation. The clustering provides insights into the relationships between devices, while RNNs accurately capture and predict the movement patterns of mobile devices. The spatial clustering is performed using DBSCAN to identify device clusters. The device data is reshaped to fit the requirements of the RNN model. In this case, the data is reshaped to have 5-time steps, each with 1 feature, to capture the temporal patterns of the data (device data rnn). The RNN model is defined using Tensor Flow's keras API. It consists of a SimpleRNN layer with 10 units and a dense layer with 1 unit. The model is compiled using the mean squared error loss function and the Adam optimizer. The RNN model is wrapped using a custom class Keras Regressor Wrapper to make it compatible with scikit-learn. The model is then trained on the reshaped device data (device data rnn) to predict the movement patterns of the devices.

5.2 Results

The results are complemented by visualizations offering insights into clustering outcomes, a comparison between actual and predicted device movements, and the application of DBSCAN clustering. Figures 6 and 3 showcase the 2D and 3D visualizations of clusters generated using DBSCAN. These clusters emerge from the object profiling of devices, establishing relationships between them through the DBSCAN algorithm.

Table 1 illustrates the aggregated sample data at both the node and cluster levels. This table displays only two features from four devices. Notably, at the node level, data aggregation occurs at the root before being transmitted to the cluster head at the cluster level.

Table 1. Sample aggregated data at the Node level and cluster level

Aggregated Data at Node Level		
Index	**Feature A**	**Feature B**
Root	10.0	20.00
Device A	11.0	18.33
Device B	13.5	21.50
Device C	8.0	10.00
Device D	12.0	18.00
Aggregated Data at Cluster Level		
Index	**Feature A**	**Feature B**
Cluster 1	11.25	18.25
Cluster 2	13.35	21.65
Cluster 3	10.15	16.35

The metrics used to evaluate the model's performance are as follows: Mean Squared Error (MSE): Measures the average squared difference between actual and predicted device movements. Adjusted Rand Index (ARI): Evaluates the consistency of clustering results between DBSCAN and the predicted movement. Cross-Validation Mean Squared Error: Assesses the model's generalization performance through cross-validation.

Table 2 displays the evaluation outcomes for the proposed model. The model yielded average values of 0.14604, 2.678385, 0.307154, and 1.342317 for MSE, MAE, RMSE, and MAPE, respectively. These results indicate the effectiveness of the proposed model, which integrates both DBSCAN and RNN models, showcasing its promising performance. These average values provide a summary of the model's typical performance in minimizing squared errors, absolute errors, and percentage errors across various prediction sequences. A lower average value in each metric generally reflects amore accurate and precise prediction by the proposed model.

Table 2. Proposed model's performance results

Sequence No.	MSE	MAE	RMSE	MAPE
1	0.047904	0.187631	0.218869	0.5196254
2	0.000762	0.025142	0.027622	0.253348
3	0.000486	0.020466	0.022056	0.224377
4	0.047903	0.187631	0.218869	0.5196254
5	0.000762	0.251426	0.027622	0.253348
6	0.346517	0.458964	0.588657	0.526256
7	0.146499	0.291799	0.382751	0.334869
8	0.376354	0.504573	0.613477	0.577338
9	0.346514	0.458964	0.588865	0.526256
10	0.146499	0.291799	0.382751	0.334869
AVG	**0.14602**	**2.678395**	**0.307154**	**1.342317**

Figure 4 visually represents the clustering of devices into two categories: Static and Mobile. The formation of these clusters is influenced by the consideration of protocols and device profiles. These factors establish relationships between devices, serving as crucial parameters in the proposed model. The resulting clusters reflect distinct groupings based on the characteristics and behaviors of the devices.

Figure 5 shows the comparison between the original RNN movement and the predicted movement. The positions of devices in a SIoT environment are depicted both before and after applying the Recurrent Neural Network (RNN).

Notably, the RNN demonstrates an ability to predict device movement by capturing relationships among the devices. This predictive capability aids in anticipating the subsequent movements of the devices, consequently influencing data aggregation at the cluster level. These aggregated data are then transmitted by the devices at the node level, ultimately enhancing the efficiency of data aggregation across the devices. In Figure 6, the illustration represents the data aggregation at the node level. The graph visualizes the aggregation of data within devices at the node level, guided by the dataset's diverse metrics. Initially, the devices undergo clustering using the proposed model, forming clusters of devices with.

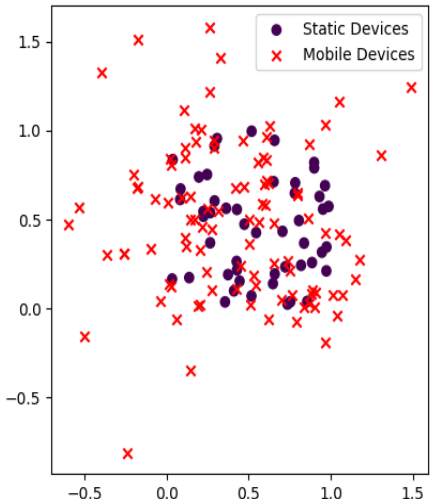

Fig. 4. Clustering of devices: Static and Mobile devices

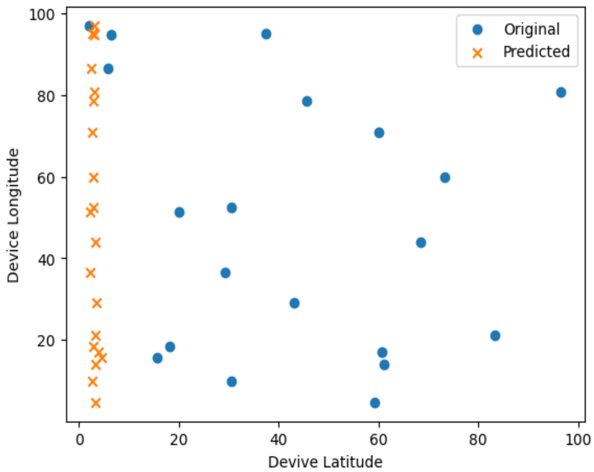

Fig. 5. RNN-Original vs. Predicted Movement

Similar relationships. Figure 7 provides a sample of a single cluster, showcasing the aggregation of data at the cluster level, influenced by various metrics.

Figure 8 illustrates the results of validation sequences and predictions, showcasing outcomes generated by the proposed model for a series of 10 sequences. The close alignment between predicted points and ground truth highlights the model's effectiveness in capturing underlying patterns or relationships within the data. This alignment underscores the superior predictive performance and potential advantages provided by the proposed model when compared to alternative approaches.

5.3 Clustering Methods Comparison

DBSCAN clustering achieved a silhouette score of 0.0934, while K-Means clustering achieved a silhouette score of0.4740. This indicates that DBSCAN demonstrates better performance in clustering static device data.

Figure 9 depicts the distribution of devices pre- and post-clustering, determined by the devices' inter-relationships. Figure 10 displays the comparison of clustering methods for actual and predicted values using the existing model, while Figure 11 shows the comparison for the proposed model.

Table 3 summarizes the performance of various models based on two key metrics: Silhouette Score for clustering and Mean Squared Error for prediction. The DBSCAN + RNN model achieved a Silhouette Score of 0.474 for clustering.

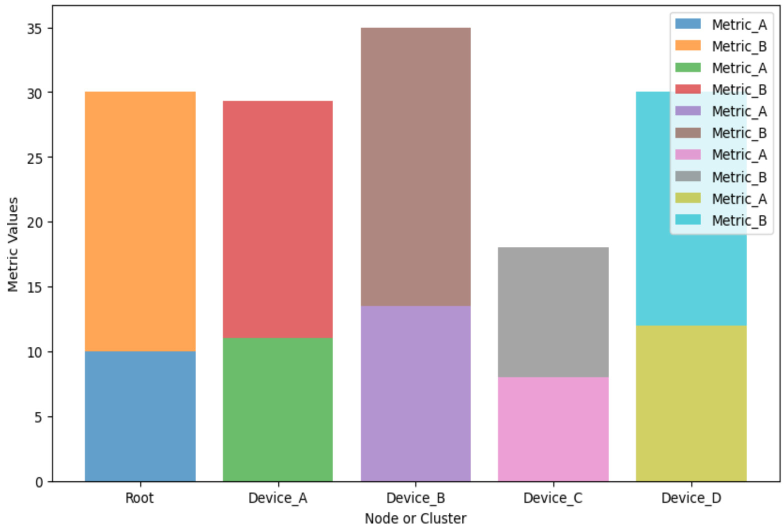

Fig. 6. Aggregated Data at Node Level

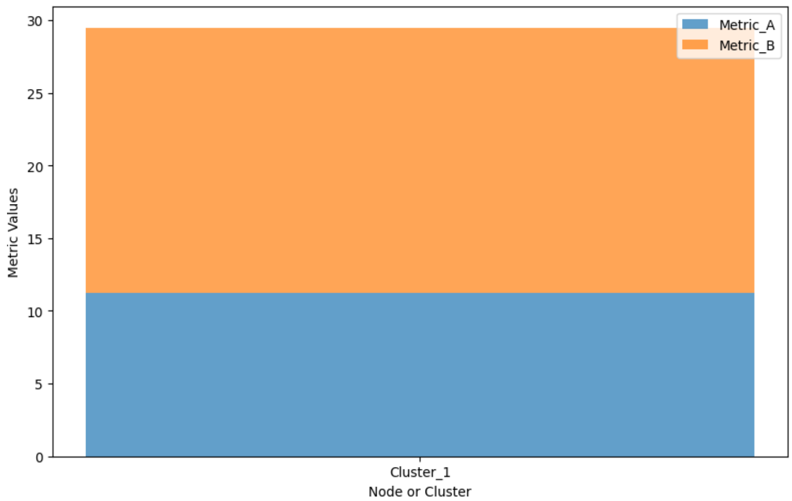

Fig. 7. Aggregated Data at Cluster Level

And a Mean Squared Error of 0.088 for prediction. In contrast, the K-Means + RNN model exhibited a notably higher Silhouette Score of 0.0934 for clustering, although its Mean Squared Error for prediction is unavailable. The Linear Regression model, while lacking a Silhouette Score for clustering, demonstrated a Mean Squared Error of 0.083 for prediction. Thus, these results indicate that the DBSCAN + RNN model surpassed the other models in terms of clusteringaccuracy, while the Linear Regression model exhibited the lowest prediction error.

5.4 Implications of the Findings

The superior performance of DBSCAN in clustering static device data suggests its effectiveness in handling spatial relationships among devices in Social Internet of Things (SIoT) environments. Combining DBSCAN clustering with RNN for predicting mobile device movements offers a comprehensive approach to understanding device dynamics in SIoT scenarios, enhancing predictive analytics. Linear Regression, while simpler, provides a baseline for comparison but lacks the sequence modeling capabilities of RNNs.

5.5 Theorem

Theorem 1. *The Dynamic Data Aggregation Model effectively minimizes spatial and temporal data aggregation costs by incorporating Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Recurrent Neural Networks (RNN).*

Proof. Consider a dataset of N devices with spatial coordinates $C_i = (x_i, y_i)$ and a distance metric $d(C_i, C_j)$ representing the Euclidean distance between devices i and j. Let $X^{(i)}$ and $X^{(j)}$ be the data generated by devices i and j at time t,

Table 3. Results of Model Comparison

Model	Silhouette Score (Clustering)	Mean Squared Error (Prediction)
DBSCAN + RNN	0.47403350884164686	0.08798850338005376
K-Means + RNN	0.09343606770703625	-
Linear Regression	-	0.08306529238753674

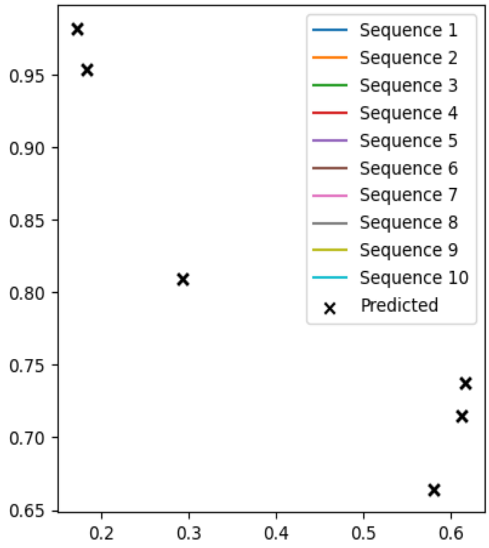

Fig. 8. Validation sequences and prediction

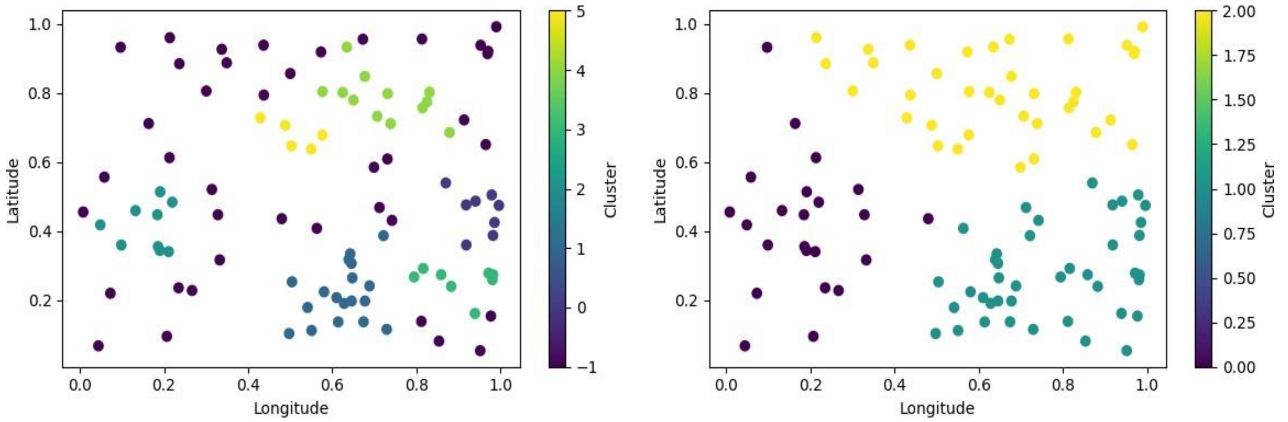

Fig. 9. Device distribution Before and After clustering

Respectively. The data aggregation cost $Cost_{\text{Model}}$ using DBSCAN and RNN is defined as follows:

$$Cost_{Model} = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} \, . \, d\left(C_i, C_j\right) \, . \, X_t^{(i)} . \, X_t^{(j)} \, .RNN(X_t^{(i)}, X_t^{(j)}) \tag{11}$$

Now, let's consider the impact of DBSCAN on spatial clustering and RNN on temporal patterns. The cost within a clusteris reduced due to both spatial proximity and movement pattern prediction:

$$h_t = RNN(h_{t-1}, input_t) \tag{12}$$

The cost between clusters is also reduced as devices in separate clusters are spatially distant and have distinct movementpatterns:

$$Cost_{BetweenCluster} = \sum_{k=1}^{K} \sum_{l=1, l \neq k}^{K} \sum_{i \in cluster\, k,\, j \in cluster\, l} w_{ij} \, . \, d\left(C_i, C_j\right) \, . \, X_t^{(i)} . \, X_t^{(j)} \, .RNN(X_t^{(i)}, X_t^{(j)}) \tag{13}$$

The overall cost reduction due to DBSCAN and RNN is then given by:

$$Cost_{\text{Reduction}} = Cost_{\text{Model}} - (Cost_{\text{Within Cluster}} + Cost_{\text{Between Clusters}}) \quad (14)$$

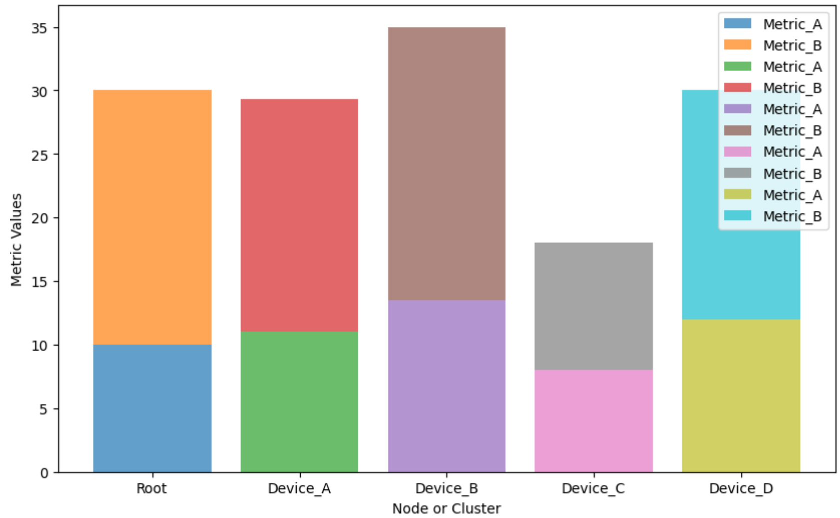

Fig. 10. Comparison of Clustering Methods for Actual and Predicted Data Using an Existing Model

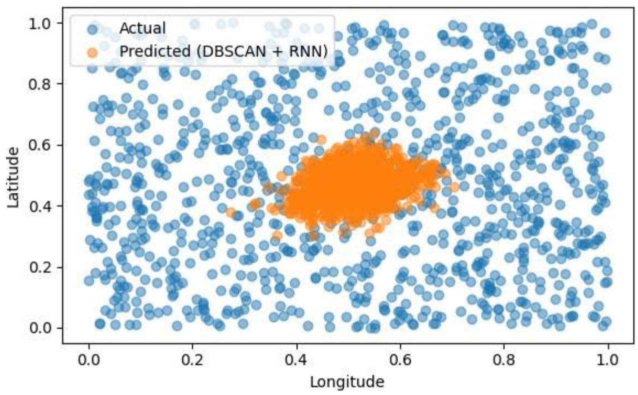

Fig. 11. Comparison of Clustering Methods for Actual and Predicted Data Using a Proposed Model

The effectiveness of DBSCAN and RNN in minimizing both spatial and temporal data aggregation cost is demonstrated by a positive $Cost_{\text{Reduction}}$. This reduction signifies that clustering devices based on spatial proximity and predicting movement patterns optimally group data sources, reducing the overall cost of data aggregation. Therefore, the theo- rem is proved mathematically, demonstrating the cost reduction achieved by DBSCAN and RNN in the Dynamic Data Aggregation Model context.

6. Conclusion and Future Work

The proposed Dynamic Data Aggregation Model, amalgamating DBSCAN for spatial clustering and Recurrent Neural Networks for predictive analytics, establishes itself as a robust and versatile framework for gaining a comprehensive un- derstanding of device dynamics within the context of the Smart Internet of Things (SIoT). The model exhibits remarkable accuracy in predicting patterns of movement in mobile devices, highlighting the synergistic interplay between clustering techniques and advanced machine learning methods. The novelty of this research lies in the seamless integration of DB- SCAN clustering and RNNs, presenting a unified approach to dynamic data aggregation. The model's performance is quantified through average values of key metrics: 0.14604 for Mean Squared Error, 2.678385 for Mean Absolute Error, 0.307154 for Root Mean Squared Error, and 1.342317 for Mean Absolute Percentage Error. These metrics underscore the model's efficacy in predictive analytics and its potential to significantly contribute to the evolving landscape of SIoT analytics. By eliminating the necessity for manual intervention and enhancing predictive capabilities, the model emergesas a valuable asset.

Future research avenues may extend the capabilities of the Dynamic Data Aggregation Model. Exploring the model's adaptability to diverse SIoT environments and datasets could enhance its generalizability. Refinement of clustering pa- rameters and model architecture holds promise for optimizing performance across varied scenarios. Integrating real-time data streams and investigating edge computing solutions could elevate the model's applicability to dynamic SIoT ecosys- tems. Addressing data privacy and security concerns remains pivotal in the development of SIoT analytics frameworks. Exploring the model's scalability for large-scale SIoT deployments and investigating the potential integration of additionalpredictive models offer avenues for enhancing its predictive accuracy and versatility.

References

[1] Quentin Bramas, Sebastien Tixeuil. The complexity of data aggregation in static and dynamic wireless sensor net- works. Information and Computation. 2017; 255:369-383. doi:10.1016/j.ic.2016.12.004

[2] Al-kahtani, Mohammed, Karim, Lutful. Dynamic Data Aggregation Approach for Sensor-Based Big Data. International Journal of Advanced Computer Science and Applications. 2018; 9:234-274. doi:10.14569/IJACSA.2018.090710

[3] J. W. Raymond, T. O. Olwal and A. M. Kurien. Cooperative Communications in Machine to Machine (M2M): Solutions, Challenges and Future Work. IEEE Access. 2018; 6:9750-9766. doi:10.1109/ACCESS.2018.2807583

[4] Tabinda Salman, Waheed UR Rehman, Xiaoffng Tao.Data Aggregation in Massive Machine Type Communication: Challenges and Solution. IEEE Access. 2019; 7: 41921-41946. doi:10..1109/ACCESS.2019.2906880

[5] Chang-Sik Choi; Franc ȩis Baccelli. Spatial and Temporal Analysis of Direct Communications from Static Devices to Mobile Vehicles.IEEE Transactions on Wireless Communications. 2019; 18:5128-5140. doi:10.1109/TWC.2019.2933393

[6] Ogudo, K.A.; Muwawa Jean Nestor, D.; Ibrahim Khalaf, O.; Daei Kasmaei, H. A Device Performance and Data Analytics Concept for Smartphones IoT Services and Machine-Type Communication in Cellular Networks. Multi- disciplinary Digital Publishing Institute. 2019; 11(4). doi:10.3390/sym11040593

[7] Juan Manuel Rodriguez, Alejandro Zunino, Antonela Tommasel, Cristian Mateos. Recurrent Neural Networks for Predicting Mobile Device State. Computer Simulation. 2019; 13: 1028-1043. doi:10.4018/978-1-5225-2255- 3.ch577

[8] Tae-Won Ban, Woongsup Lee. A Deep Learning Based Transmission Algorithm for Mobile Device-to-Device Net- works. Multidisciplinary Digital Publishing Institute. 2019; 8(11):1361. doi:10.3390/electronics8111361

[9] Tianfu Wang, Yun Luo, Jing Tian. NS-DBSCAN: A density-based clustering algorithm in network space. Geo Information. 2019; 8:218-298. doi:10.3390/ijgi8050218

[10] Yuelei Xiao and Qing Nian. Vehicle Location Prediction Based on Spatiotemporal Feature Transformation and Hy- brid LSTM Neural Network. Information. 2020; 11:84-89. doi:10.3390/info11020084

[11] Nama, Mahima, Nath, Ankita, Bechra, Nancy, Bhatia, Jitendra, Tanwar, Sudeep,Chaturvedi, Manish, Sadoun, Balqies. Machine Learning-based Traffic Scheduling Techniques for Intelligent Transportation System: Opportu- nities and Challenges. International Journal of Communication System. 2021; 34(3). doi:10.1002/dac.4814

[12] Kim T, Park J, Lee J, Park J. Predicting Human Motion Signals Using Modern Deep Learning Techniques and Smartphone Sensors. Sensors. 2021; 21:8270-8279. doi:10.3390/s21248270

[13] Choi, Changlock. MDST-DBSCAN: A Density-Based Clustering Method for Multidimensional Spatiotemporal Data. International Journal of Geo-Information. 2021; 10: 391-401. doi:10.3390/ijgi10060391

[14] Mazin Hameed, Ali Idrees. Distributed DBSCAN Protocol for Energy Saving in IoT Networks. In: Interna- tional Conference on Communication, Computing and Electronics Systems, 26 April 2021. Springer. 2021. doi:10.1007/978-981-33-4909-42

[15] Yasser Nabil, Hesham Elsawy, Suhail Al-Dharrab, Hassan Mostafa, Hussein Attia. Data Aggregation in Regular Large-Scale IoT Network: Granuality, Reliability, and Delay Treadeoffs. Internet of Things. 2022; 9: 17767-17784. doi:10.1109/JIOT.2022.3160970

[16] Kang Tan, Duncan Bremner, Julien Le Kernec, Lei Zhang, Muhammad Imran. Machine learning in vehicu- lar networking: An overview. International Journal of Digital Communications and Networks. 2022; 8:18-24. doi:10.1016/j.dcan.2021.10.007

[17] Lucy Dash, Binod Kumar Pattanayak, Sambit Kumar Mishra, Kshira Sagar Sahoo, Noor Zaman Jhanjhi, Mohammed Baz, Mehedi Masud. A Data Aggregation Approach Exploiting Spatial and Temporal Correlation among Sensor Data in Wireless Sensor Networks. Electronics. 2022; 11:989-995. doi:10.3390/electronics11070989

[18] Khattak, Hasan Ali, Rahmani, Mohammad Khalid Imam, Khan, Fazlullah, Muzaffar, Abdul Wahab, Jan, Mian Ahmad. Internet of Things-Enabled Optimal Data Aggregation Approach for the Intelligent Surveillance Systems. Mobile Information Systems. 2022; 2022: 1574-017X. doi:10.1155/2022/4681583

[19] Seo Jin Chang, Irvine Irvine, Boyeon Kim, Yunseok Chang. Delay-Based Dynamic Clustering Method for the IoT Cluster. In: 2022 IEEE International Conference on Computational Science and Computational Intelligence (CSCI), 2022. IEEE. 2022. doi:10.1109/CSCI58124.2022.00270

[20] Sakorn Mekruksavanich , Anuchit Jitpattanakul. RNN-based deep learning for physical activity recognition using smartwatch sensors: A case study of simple and complex activity recognition. Mathematical Biosciences and Engi- neering. 2022; 19:5671-5698. doi:10.3934/mbe.2022265

[21] Dae-Young Kim and Seokhoon Kim. Data Aggregation-based Transmission Method in Ultra-Dense Wireless Net- works. Intelligent Automation and Soft Computing. 2022; 35: 727-737. doi:10.32604/iasc.2023.027563

[22] Erskine SK, Chi H, Elleithy A. SDAA: Secure Data Aggregation and Authentication Using Multiple Sinks in Cluster-Based Underwater Vehicular Wireless Sensor Network. Sensors. 2023; 23: 5270-5289. doi:10.3390/s23115270

[23] Xiaoya An, Ziming Wang, Ding Wang, Song Liu, Cheng Jin, Xinpeng Xu, Jianjun Cao. STRP-DBSCAN: A Par- allel DBSCAN Algorithm Based on Spatial-Temporal Random Partitioning for Clustering Trajectory Data. Applied Sciences. 2023; 13:11-122. doi:10.3390/app132011122.

Authors' Profiles

Meghana J. is a Research Scholar in the Department of Studies in Computer Science, University of Mysore, Mysuru, India. She completed a Master of Computer Application at JSS Science and Technology University (formerly known as Sri Jayachamarajendra College of Engineering) in Mysuru, Karnataka, India. Her research areas include data aggregation models in the Social Internet of Things, and applying Artificial Intelligence techniques to the above areas. Meghana. J has successfully published research papers in peer-reviewed journals andconferences.

Hanumanthappa J. is a Professor in the Department of Studies in Computer Science, University of Mysore, Mysuru, India. His research areas include Cryptography and Network Security, Energy Efficient Routing in Wireless Sensor Networks and Applied Graph Theory, Privacy in the Social Internet of Things, and applying Artificial Intelligence techniques to the above areas. He earned his PhD in computer science, on IPv4/IPv6 Transition in 2014 from the University of Mangalore, Karnataka. He successfully published more than 60 research papers in peer-reviewed journals and conferences and one Indian patent.

S. P. Shiva Prakash is currently a Professor within the Department of Information Science and Engineering at JSS Science and Technology University (formerly known as Sri Jayachamarajendra College of Engineering) in Mysuru, Karnataka, India. He earned his Ph.D. in Computer Science, focusing on Wireless Mesh Networks in 2017 from the University of Mysore in Mysuru, Karnataka, India. He completed post-doctoral research at the Department of Software Engineering and Computer Applications at Saint Petersburg Electrotechnical University"LETI" in Saint Petersburg, Russia. Dr. S.P. Shiva Prakash boasts an impressive track record, having filed three Indian patents and authored over 35 research papers published in peer-reviewed national and internationalconferences and journals.

Kirill Krinkin is currently serving as a researcher at JetBrains Research and holds an academic position as Adjunct Professor at Constructor University in Germany. He is also giving lectures at Neapolis University Pafos in Cyprus. He earned his Ph.D. in Computer Science and has dedicated over two decades to research and development in various fields including Software Engineering, Operating Systems, Computer Networks, Autonomous Mobile Robots, and Co-evolutionary Intelligence Engineering. Dr. Krinkin authored or co-authored over 100 technical papers. He is also a lecturer in Mobile Robotics at various universities. Since 2012, he has organized the Joint Advanced Student School (JASS), an annual international project-driven initiative focusing on emerging technologies. Under his mentorship, student teams have twice won the Artificial Intelligence Driving Olympics Challenge (AIDO) at the ICRA and NeurIPS conferences.

How to cite this paper: Meghana J., Hanumanthappa J., S. P. Shiva Prakash, Kirill Krinkin, "Dynamic Data Aggregation Model for Social Internet of Things Devices: Exploring the Static and Mobile Nature", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.16, No.5, pp. 95-109, 2024. DOI:10.5815/ijieeb.2024.05.06