IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, June 2026
Impact Factor: 8.2

Data Mining-Based Optimization of Data Transfer Efficiency in Internet of Things Networks

Shwetha H L[1] and Dr. Deepak Kumar Sharma[2]
[1]Research Scholar, [2]Professor
Department of Computer Science
Sunrise University, Alwar, Rajasthan

Abstract: *The Internet of Things (IoT) has emerged as a fundamental technological paradigm for connecting sensors, devices, machines, and intelligent systems through communication networks. The rapid growth in the number of IoT devices has resulted in massive volumes of heterogeneous data being generated and transferred continuously across resource-constrained networks. Consequently, efficient data transfer has become an important requirement for reducing communication overhead, energy consumption, latency, bandwidth utilization, and packet loss. Data mining techniques provide effective mechanisms for discovering hidden patterns, correlations, redundancies, and transmission behaviours within IoT data. This review examines the application of data mining-based approaches for optimizing data transfer efficiency in IoT networks.*

The discussion focuses on data preprocessing, feature selection, clustering, classification, association analysis, predictive analytics, and intelligent routing as mechanisms for improving communication performance. Data mining can support efficient data aggregation by identifying redundant information and selecting representative data before transmission. Clustering methods can organize sensor nodes and reduce unnecessary communication, while predictive models can estimate traffic patterns and transmission requirements.

Keywords: Internet of Things, Data Mining, Clustering, Intelligent Routing, Energy Efficiency, Machine Learning, IoT Networks, Network Optimization.

I. INTRODUCTION

The Internet of Things represents a rapidly developing network environment in which physical objects are equipped with sensing, processing, communication, and identification capabilities. IoT networks are increasingly deployed in smart cities, healthcare, industrial automation, agriculture, transportation, environmental monitoring, smart homes, and energy management. These applications generate continuous streams of data that must be transmitted efficiently between sensing devices, gateways, edge nodes, cloud platforms, and end users. The large-scale deployment of IoT devices creates significant communication challenges because many devices have limited battery capacity, processing power, memory, and bandwidth. Therefore, optimizing data transfer is essential for improving network lifetime and application performance.

Data transfer efficiency refers to the ability of an IoT network to deliver useful information while minimizing unnecessary communication, energy consumption, latency, congestion, and packet loss. Conventional transmission approaches frequently send large quantities of raw sensor data even when substantial portions of the information are duplicated or have limited relevance. Such transmission increases network traffic and consumes valuable energy. Data mining provides a systematic means of examining large IoT datasets to identify useful patterns and relationships that can support intelligent transmission decisions.

Data mining-based optimization can be applied at multiple stages of IoT communication. During preprocessing, noisy, incomplete, and redundant data can be removed. Feature selection can reduce the dimensionality of transmitted information, while clustering can group similar observations or sensor nodes. Classification can distinguish important information from less significant data, and predictive analytics can anticipate future network traffic or sensor values. These capabilities allow IoT systems to transmit only the information that is necessary for decision-making.

The integration of data mining with edge and fog computing further strengthens this approach. Instead of transmitting all raw data to centralized cloud servers, computational operations can be performed closer to data sources. This reduces transmission distance and latency while conserving bandwidth. Intelligent algorithms can also continuously learn from network conditions and adjust transmission strategies. Thus, data mining is increasingly relevant for building adaptive IoT networks capable of handling high data volumes and changing communication requirements.

II. DATA MINING TECHNIQUES FOR IOT DATA TRANSFER OPTIMIZATION

Data mining encompasses several computational techniques that can be adapted to IoT communication environments. Among the most important are clustering, classification, association rule mining, regression, anomaly detection, and predictive analytics. Each technique can address a different aspect of data transfer optimization.

Clustering is particularly useful in wireless sensor and IoT networks because similar sensor nodes or observations can be grouped together. Cluster-based communication allows a selected cluster head or gateway to aggregate information before forwarding it to another network component. This reduces the number of individual transmissions and can significantly lower communication energy consumption. Clustering algorithms such as K-means, hierarchical clustering, density-based methods, and fuzzy clustering can be adapted according to network structure and application requirements.

Classification techniques can categorize IoT data according to importance, urgency, type, or transmission priority. For example, healthcare applications may classify physiological information into normal and critical categories, allowing urgent data to receive higher transmission priority. Decision trees, support vector machines, random forests, and neural networks can be used for such classification tasks.

Association rule mining can identify relationships between sensor observations. When multiple sensors frequently produce correlated information, the system can avoid transmitting repetitive measurements. Regression and predictive mining techniques can estimate future sensor values from historical observations. If the predicted value remains within an acceptable range, an IoT device may transmit only when a significant deviation occurs. This event-driven strategy can substantially reduce unnecessary transmissions.

Anomaly detection is another important application. Data mining models can identify unusual traffic patterns, sensor readings, or communication behaviour. Detecting anomalies helps prevent unnecessary transmission of abnormal or corrupted information and can also support network security. Consequently, data mining should not be viewed solely as a data analysis mechanism; it can function as an intelligent decision layer between sensing and communication processes.

III. DATA AGGREGATION, COMPRESSION AND REDUNDANCY REDUCTION

Data aggregation is one of the most direct methods for improving IoT data transfer efficiency. Sensor networks often contain multiple devices monitoring the same physical phenomenon, resulting in highly correlated observations. Transmitting every observation independently increases bandwidth consumption and energy expenditure. Data mining can identify similarities among observations and determine which information should be aggregated before transmission.

Clustering-based aggregation organizes nodes or data points into groups and enables representative information to be transmitted instead of every individual measurement. Statistical techniques can calculate averages, medians, minimums, maximums, or other representative values depending on application requirements. More advanced mining methods can identify temporal and spatial correlations, enabling systems to suppress data that provides little additional information.

Data compression is another important mechanism. Data mining can identify frequently occurring patterns and structures that can be represented using fewer bits. Feature selection can further remove attributes that do not contribute substantially to the intended application. Dimensionality-reduction methods such as principal component analysis can transform large datasets into smaller representations while retaining important information.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, June 2026
Impact Factor: 8.2

Prediction-based transmission provides another strategy for redundancy reduction. A gateway or edge device can predict expected sensor readings using historical data. The sensor transmits a new measurement only when its actual value differs from the predicted value beyond a predefined threshold. Such approaches can considerably decrease communication frequency in relatively stable environments. However, excessive compression or aggressive redundancy removal may result in information loss. Therefore, optimization methods must balance communication savings against data accuracy and application requirements.

IV. INTELLIGENT ROUTING AND RESOURCE OPTIMIZATION

Efficient routing is essential for IoT networks because communication paths directly affect energy consumption, latency, reliability, and network lifetime. Traditional routing protocols may not perform optimally when network topology, traffic volume, node availability, or channel conditions change dynamically. Data mining can analyze historical and real-time network information to identify efficient routing strategies.

Predictive data mining can estimate congestion levels and determine the likelihood of packet loss on different communication paths. Routing algorithms can then select paths with lower congestion and better reliability. Classification models can also categorize network conditions and trigger different routing strategies for different situations. For example, a network may prioritize energy-efficient routes under normal conditions but select low-latency routes when time-sensitive information is detected.

Clustering can support hierarchical routing by creating groups of sensor nodes and reducing long-distance communication. The selection of cluster heads can be optimized according to residual energy, distance, communication quality, workload, and node importance. Data mining models can analyze these parameters and identify nodes that are suitable for coordinating local communication.

Resource optimization also includes bandwidth allocation, transmission scheduling, and energy management. Predictive models can estimate future traffic demand and allow network resources to be allocated before congestion occurs. Reinforcement-learning and machine-learning-assisted approaches can further adapt resource allocation according to changing conditions. The combination of data mining and intelligent routing therefore creates opportunities for self-adaptive IoT communication architectures.

V. COMPARATIVE ANALYSIS OF DATA MINING APPROACHES

The effectiveness of a data mining technique depends on the IoT application, network architecture, data characteristics, computational resources, and optimization objective. No single technique is universally optimal. Clustering is highly useful for reducing communication among similar nodes, while classification is suitable for prioritizing data. Predictive methods are advantageous when sensor values exhibit temporal stability, whereas anomaly detection is valuable for identifying abnormal measurements and network behaviour.

Data Mining Technique	**IoT Application**	**Contribution to Data Transfer**	**Major Benefit**	**Major Limitation**
K-Means Clustering	Sensor grouping and aggregation	Reduces duplicate transmissions	Simple and computationally efficient	Requires predefined number of clusters
Hierarchical Clustering	Multi-level IoT networks	Supports structured data aggregation	Suitable for hierarchical networks	Can have high computational cost
Classification	Data prioritization	Separates critical and non-critical data	Enables priority-based transmission	Requires suitable training data
Decision Tree	Routing and traffic classification	Supports transmission decisions	Easy to interpret	May become complex for large datasets
Random Forest	Network-state prediction	Improves routing and resource decisions	Good predictive performance	Higher processing requirements

Support Vector Machine	Traffic and anomaly classification	Identifies transmission patterns	Effective with high-dimensional data	Computationally demanding for large datasets
Association Rule Mining	Sensor correlation analysis	Reduces redundant information	Identifies relationships among data	Can generate many irrelevant rules
Regression	Sensor-value prediction	Enables event-driven transmission	Reduces transmission frequency	Performance depends on data stability
Principal Component Analysis	Dimensionality reduction	Decreases data volume	Retains important information	Some information may be lost
Anomaly Detection	Fault and abnormal-data identification	Prevents unnecessary transmission	Improves data quality and reliability	False positives may occur
Neural Networks	Traffic and sensor prediction	Supports adaptive transmission	Handles complex nonlinear relationships	Requires greater computational resources
Deep Learning	Large-scale IoT analytics	Enables advanced prediction and optimization	High learning capability	High energy and processing requirements
Reinforcement Learning	Adaptive routing	Optimizes routes dynamically	Responds to changing environments	Training can be time-consuming

The comparative analysis indicates that lightweight data mining methods remain important for resource-constrained IoT devices. Complex machine-learning models may provide higher predictive accuracy but can impose substantial computational and energy costs. Edge computing can address this limitation by moving computationally intensive operations away from low-power sensor nodes. Future systems are therefore likely to employ hybrid architectures in which lightweight processing occurs at sensors while advanced data mining is performed at edge gateways.

VI. CHALLENGES, RESEARCH GAPS AND FUTURE DIRECTIONS

Despite the potential of data mining-based optimization, several challenges remain. IoT networks are highly heterogeneous, containing devices with different communication technologies, computational capabilities, sensing characteristics, and energy constraints. A mining algorithm designed for one network environment may not perform effectively in another. Dynamic topology and changing traffic conditions further complicate optimization because models must continuously adapt to new information.

Computational complexity is another major concern. Advanced data mining and deep-learning models can require considerable processing power and memory, which may be unavailable on small IoT devices. Distributed and edge-based data mining can reduce this burden, but it introduces additional requirements for model synchronization, communication, and resource management.

Data privacy and security also require greater attention. IoT data may contain sensitive information, and transmitting data to centralized servers can create privacy risks. Privacy-preserving data mining, federated learning, secure aggregation, and decentralized analytics represent promising directions for future research. Furthermore, data mining models themselves can be vulnerable to noisy, manipulated, or adversarial data.

Future research should focus on lightweight and adaptive data mining algorithms specifically designed for IoT environments. Explainable models can improve the transparency of automated transmission decisions, while federated approaches can enable collaborative learning without requiring raw data to be centrally collected. The integration of data mining with edge computing, software-defined networking, digital twins, and intelligent communication protocols also deserves further investigation. Energy-aware model selection, real-time predictive transmission, semantic communication, and application-aware data reduction are additional promising research areas.

Overall, data mining-based optimization offers an effective pathway for improving IoT data transfer efficiency. By identifying redundancy, predicting data and traffic behaviour, optimizing routes, prioritizing information, and supporting intelligent resource allocation, data mining can reduce communication overhead while maintaining acceptable data quality. Future IoT networks will increasingly require integrated approaches that combine data mining, artificial intelligence, edge computing, and adaptive communication mechanisms to achieve scalable, reliable, secure, and energy-efficient data transmission.

REFERENCES

1. Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., & Ayyash, M. (2015). Internet of Things: A survey on enabling technologies, protocols, and applications. *IEEE Communications Surveys & Tutorials, 17*(4), 2347–2376.
2. Atzori, L., Iera, A., & Morabito, G. (2017). Understanding the Internet of Things: Definition, potentials, and societal role. *Computer Networks, 136*, 1–7.
3. Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems, 29*(7), 1645–1660.
4. Xu, L. D., He, W., & Li, S. (2014). Internet of Things in industries: A survey. *IEEE Transactions on Industrial Informatics, 10*(4), 2233–2243.
5. Zanella, A., Bui, N., Castellani, A., Vangelista, L., & Zorzi, M. (2014). Internet of Things for smart cities. *IEEE Internet of Things Journal, 1*(1), 22–32.
6. Bonomi, F., Milito, R., Zhu, J., & Addepalli, S. (2012). Fog computing and its role in the Internet of Things. *Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing*, 13–16.
7. Sicari, S., Rizzardi, A., Grieco, L. A., & Coen-Porisini, A. (2015). Security, privacy and trust in Internet of Things: The road ahead. *Computer Networks, 76*, 146–164.
8. Miorandi, D., Sicari, S., De Pellegrini, F., & Chlamtac, I. (2012). Internet of Things: Vision, applications and research challenges. *Ad Hoc Networks, 10*(7), 1497–1516.
9. Al-Fuqaha, A., Khreishah, A., Guizani, M., Rayes, A., & Mohammadi, M. (2015). Data-centric design of wireless sensor networks: A survey. *IEEE Communications Surveys & Tutorials, 17*(2), 1087–1111.
10. Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys, 46*(4), 1–37.