

PropMat-DAE: An Avionics System Fault Diagnosis Algorithm based on Graph Anomaly Detection

Tianyi Li¹ and Lisong Wang¹

¹ School of computer science and technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China
tianyi@nuaa.edu.cn

Abstract. At present, in the field of sensor anomaly detection in avionics systems, most studies only consider the temporal anomalies of individual sensors, and rarely consider the hidden spatial position relationships between sensors, so the spatiotemporal anomalies of nodes cannot be fully considered. This article proposes the PropMat-DAE framework, which is an algorithm for diagnosing anomalies from both temporal and spatial perspectives. Based on sensor timing data, the graph structure learning algorithm is used to calculate the implicit spatial position relationship between nodes, and statistical methods are used to compress the original temporal features of nodes, thereby constructing a spatiotemporal attribute graph structure. Design a global attention mechanism to aggregate information from all nodes to the target node. Using matrix factorization and deep autoencoder, the reconstruction error of node attributes and structure is calculated as the spatiotemporal anomaly score of the node, and the top-k node with the highest anomaly value is selected as the fault node. After experimental verification, the algorithm proposed in this article has better accuracy than the existing 14 graph anomaly detection algorithms on small datasets. In the comparative experiment, the attention mechanism and feature compression method designed in this article have significant effects in the field of avionics.

Keywords: Fault diagnosis, GNN, Autoencoder, Matrix Decomposition.

1 Introduction

It is widely acknowledged that avionics systems play a vital role in modern aircraft. Therefore, sensor fault detection has become a popular research area. Figure 1 shows the network topology structure of the various components and their composition in the spacecraft turbine engine. The main purpose of this article is to construct a node topology structure based on sensor timing data, and then calculate node anomaly scores from both spatiotemporal and temporal perspectives, in order to diagnose faulty nodes.

In the past few decades, a large amount of research has been conducted to detect sensor failures. Traditional fault detection methods are usually based on fault trees,

models [2,3], and data-driven [10]. There are also some studies based on genetic algorithms[10,11] and particle swarm optimization[12] for fault diagnosis. However, these methods require accurate mathematical models of the process, and cannot take advantage of the hidden position relationships between sensors in non-Euclidian space, so they are just suitable for small systems with fewer inputs and clear mathematical models. However, in the presence of unmodeled interference and uncertainty, these performance is greatly affected. In recent years, graph neural networks have emerged, and in the real world, failures can alter the running state of a process and lead to changes in the

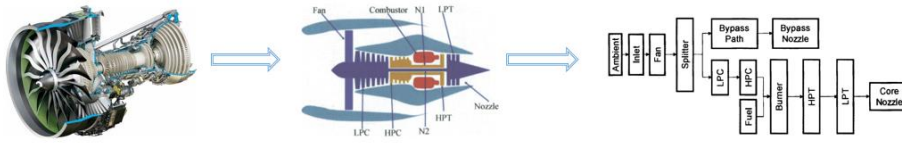


Fig.1 Schematic diagram of spacecraft turbine engine structure

dependencies between measurements. Based on this understanding, the researchers began to explore GNN for fault diagnosis tasks. A large number of researchers start from the perspective of the system topology diagram and use the method based on the structure diagram[4,8]to diagnose whether the sensor has a fault.

Although deep neural network models[7] show excellent performance in the field of fault detection, we still need to face four key problems. First, most of the current methods mainly focus on anomaly prediction or classification of tabular data to calculate anomaly scores for fault diagnosis, but do not fully consider the correlation hidden in the Non-Euclidean space between sensor data. Secondly, GNN method usually needs to obtain the connection topology of sensor network accurately. But in practical application, it is often quite difficult to obtain this information. Third, current fault detection methods based on GNN mostly focus on regression tasks of temporal anomalies and detect sensor anomalies from the perspective of root causes, but do not achieve direct sensor anomaly detection in one step. Finally, in the field of graph anomaly detection, the abnormal scores of nodes in the graph are often detected by autoencoder or matrix decomposition, but they don't take the different dimension between sensor network topology and attribute into account.

In order to overcome these limitations, this paper constructs a sensor network topology based on sensor timing data, and proposes a new node anomaly detection algorithm based on property matrix decomposition and deep autoencoder fusion of graph structure (PropMat-DAE) to obtain node anomaly scores. We use two parallel exception score markers in PropMat-DAE for the link structure and attribute of the node, respectively. This frame can adaptively calculate anomaly scores, according to high and low dimensions data, and aggregate the structure anomaly and attribute anomaly into the final anomaly scores of nodes. The main contributions of this work are summarized as follows:

(1) Spatial relationship extraction: We adopt an algorithm named Gumbel-softmax in this study to determine whether two sensors are connected through probability calculation.

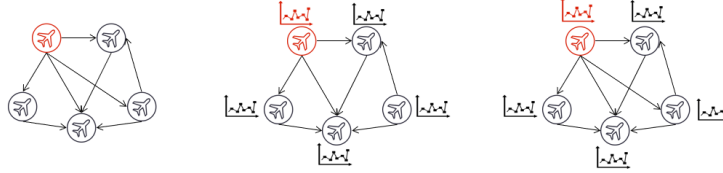


Fig. 2 Three abnormal situations in the avionics system

(2) PropMat-DAE algorithm: We introduce PropMat-DAE algorithm to solve the problem of large dimensional differences between graph structure and node attributes.

(3) Attention design: We design an all pair attention mechanism that can capture the influence of distant neighbors on the target node in a short step.

This article is mainly divided into the following parts: The second section introduces the problem to be solved in this article and how to transform it into an anomaly detection problem on the graph. The third section introduces the PropMat-DAE framework. And detailed explanations were provided from three aspects: graph structure learning, attribute anomaly detection, and structural anomaly detection. The fourth section conducts experimental evaluation using two types of spacecraft time series data, and compares it with other graph anomaly detection algorithms, attention mechanisms, and graph embedding algorithms. Section 5 provides a summary of the entire article and analyzes its limitations and future research directions.

2 Problem Formulation

We will discuss sensor anomalies that may cause malfunctions in the avionics system. The abnormal behavior of sensor topology network nodes can be captured through their link structure, attributes, or a combination of both, as shown in Figure 2. More specifically, network failures in the avionics system are mainly caused by three types of sensor anomalies. One of them is abnormal sensor interaction. In a sensor network, some nodes have far more edges connected to other nodes than others, which can be interpreted as an abnormal situation called a center node anomaly. The second is abnormal sensor timing characteristics. Under normal circumstances, attribute values may fluctuate within a certain range or change over time, but in some cases, attribute values may deviate from normal expected patterns and exhibit abnormal behavior. The third is the spatiotemporal anomaly of sensors, where nodes belong to the same community structurally but belong to different communities in terms of attribute similarity. Our goal is to learn an encoding and decoding structure that calculates the difference between sensor nodes after decompression and the original representation. Figure 4 shows the framework of the entire algorithm.

3 Solution Approach: PropMat-DAE

3.1 Gumbel-Softmax Sampling

In this section, we mainly explain how to learn the spatial connection position relationship between various sensors through time series data. Obtaining the sampling process of discrete data from sensor nodes in temporal data is non-differentiable, which leads to the inability to perform backpropagation optimization of network structures in structural learning. [6] introduced the Gumbel-Softmax distribution, it is a differentiable alternative for discrete variables in random sampling. This method enables approximate sampling from a categorical distribution, achieving differentiability in continuous variables.

For a candidate node set M , we define $V_{i,j}$ as the connectivity control variable for any pair of nodes i and j in M . The one-way probability from node i to j is defined

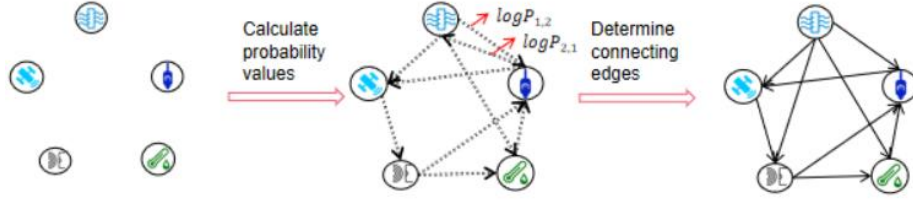


Fig.3. Assuming there are three sensors (N1,N2,N3), the dependency relationship between them is hidden. The main idea of the connection learning strategy is to use the Gumbel-Softmax sampling strategy to sample a random classification vector to determine whether any directed connections can be established between two nodes. For temperature and humidity sensors, if the value of $P_{1,2}$ is relatively large, it indicates that the temperature sensor is highly likely to point towards the humidity sensor, and vice versa.

$\{\rho_0^{i,j}, \rho_1^{i,j}\}$, where $\rho_0^{i,j} + \rho_1^{i,j} = 1$, and $\rho_1^{i,j}$ represents the flow of information between i and j , as illustrated in Figure 3. By applying the Gumbel-Max strategy, we can sample the connectivity between any i and j as shown in formula (1):

$$V_{i,j} = \operatorname{argmax}(\log \rho_a^{i,j} + g_a^{i,j}) \quad (1)$$

where g_0, g_1 is an independent identically distributed sample sampled from the standard Gumbel distribution. By plotting $u \sim \text{Uniform}(0, 1)$ and calculating $g = -\log(-\log u)$, it is easy to use inverse transform sampling for sampling. We further use the Softmax reparameterization technique (also known as the Gumbel-Softmax method) to replace the argmax operation, as the function is non differentiable. The replaced formula is shown in formula (2):

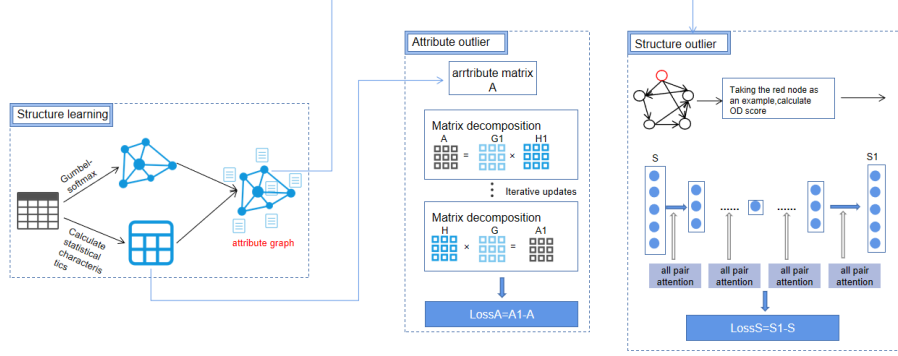


Fig.4 PropMat-DAE algorithm frame

$$V_a^{i,j} = \frac{\exp((\log \rho_a^{i,j} + g_a^{i,j})/\mu)}{\sum_{v \in \{0,1\}} \exp((\log \rho_a^{i,j} + g_a^{i,j})/\mu)} \quad (2)$$

Where $a \in \{0,1\}$ and μ is the smoothing coefficient controlling the Gumbel-Softmax distribution. When μ approaches 0, the current connection probability also approaches 0, making it identical to the one-hot encoding distribution. Due to the randomness of g and independence of ρ , we can use gradient descent to optimize the connection strategy. Compared with previous methods that select the top-k neighbors as connections, this approach reduces the subjectivity in choosing the value of k and optimizes the time complexity from $O(M^2)$ to $O(1)$, since it does not require computing the dot product between node embedding vectors.

3.2 Sensor Attribute Representation Learning

After the above steps, we can obtain the sensor attribute map of the avionics system. We believe that the temporal data of sensors can be reduced to low dimensional data through statistical feature extraction (such as mean, variance, etc). Therefore, for node attribute embedding, we choose an automatic encoder based on matrix decomposition. For the attribute graph learned before, the attribute representation of node i can be expressed by the i -th row of the attribute matrix A . Let $G' \in R^{N \times k}$ be the embedding representation of the original graph G . Thus, G'_i represents the K -dimensional representation of node i . Introduce a utility $K \times N$ dimensional matrix H to construct our loss function: $\sum_{i=1}^N \sum_{j=1}^N (A_{ij} - G'_i \cdot H_j)^2$, where $G'_i \cdot H_j$ denotes the dot product operation, and the k -th row of H is then-dimensional attribute description of the k -th sensor, where $k = 1, 2, \dots, K$.

The above loss function aims to maintain a small distance between the lower-dimensional embeddings and the original matrix. However, if there are outlier nodes in the graph, they will negatively impact the embeddings of other normal

nodes. To mitigate this negative impact in the loss function, we introduce an anomaly factor o_i , where $0 < o_i < 1$, which represents the abnormality of each sensor node's attributes. The larger o_i , the higher the likelihood of the node being abnormal, and its influence on the overall embedding should be relatively small. After introducing the anomaly factor, the loss function is as shown in formula (3):

$$L_a = \sum_{i=1}^N \sum_{j=1}^N \log\left(\frac{1}{o_i}\right) (A_{ij} - G_i \cdot H_i)^2 \quad (3)$$

In the above equation, it can be observed that when the o_i value is very high, it suppresses the value of $\sum_{j=1}^N \log\left(\frac{1}{o_i}\right) (A_{ij} - G_i \cdot H_i)^2$, thus making the contribution of that node negligible to the global context. Conversely, when o_i is very low, the node is considered significant. Therefore, the optimization of the loss function will pay more attention to those normal nodes.

In the updates of G, H within the loss function, we employ a method of controlling variables, updating only one variable at a time and alternating minimization. We need to compute the partial derivative for each variable and set

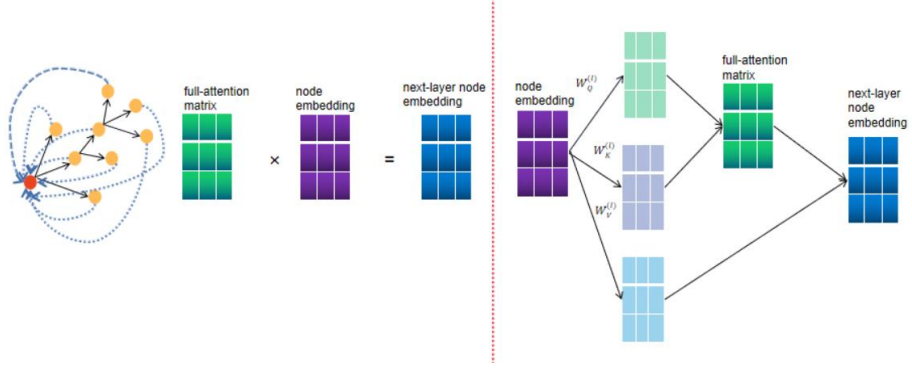


Fig.5 all pair attention graph

it to zero. For example, taking a partial derivative of the submatrix G_i of the adjacency matrix of node i and assigning it to formula(3), the result is shown in formula(4), formula(5).

$$\frac{\partial L_a}{\partial G_i} = 0 \quad (4)$$

$$\sum_{j=1}^N \log\left(\frac{1}{o_i}\right) (A_{ij} - G_i \cdot H_i) (-H_j) = 0 \quad (5)$$

Propose G_i and consider the decomposition dimension k to obtain the update formula for the decomposition sub matrix 1, as shown in formula(6).

$$G_{ik} = \log\left(\frac{1}{o_i}\right) \sum_{j=1}^N (A_{ij} - \sum_{k' \neq k} G_{ik'} H_{k'j}) H_{kj} \quad (6)$$

Similarly, we can obtain the update rule of H,as shown in formula(7).

$$H_{kj} = \frac{\sum_{i=1}^N \log(\frac{1}{\sigma_i}) (A_{ij} - \sum_{k \neq k} G_{ik} H_{kj}) G_{ik}}{\sum_{i=1}^N \log(\frac{1}{\sigma_i}) G_{ik}^2} \quad (7)$$

Then we can update o_i by(8).

$$o_i = \frac{\mu \cdot \sum_{j=1}^N (A_{ij} - G_i \cdot H_i)^2}{\sum_{i=1}^N \sum_{j=1}^N (A_{ij} - G_i \cdot H_i)^2} \quad (8)$$

3.3 Sensor Structure Characterization Learning

Due to the complex interaction between sensors, this article believes that the link structure data of sensor nodes is high-dimensional data, and adopts a deep learning encoding and decoding structure to calculate the anomaly score. The network topology of the raw data is from algorithm,so the connections between sensors may not necessarily have practical physical significance in the real world. Therefore, we introduces a global attention mechanism to dynamically update the embedded representation of each sensor node, as shown in Figure 5. We call it the diffusion coefficient $S(Z(t), t)$, it determine show information flows between instances and the direction of instance state evolution. The specification is greatly flexible, for example, a basic choice is to fix $S(Z(t), t)$ as an identity matrix, which limits feature propagation to self cycling, and the model degenerates into an MLP that independently handles all instances. We can also specify $S(Z(t), t)$ as the observed graph structure, so that the model evolves into a GCN model. In this case, the information flow is limited by adjacent nodes in the graph. In our work, we specify the form of S as shown in formula (9). Under this form, S can learn the embedding information of all layer nodes. This diffusion process can serve as an inductive bias, guiding the model to use information from other instances at each layer to learn informative instance representations.

$$S_{ij}^{(k)} = \frac{f(\|z_i^{(k)} - z_j^{(k)}\|_2^2)}{\sum_{l=1}^N f(\|z_i^{(k)} - z_j^{(k)}\|_2^2)}, 1 \leq i, j \leq N \quad (9)$$

Where $z_i^{(k)}$ represents the embedded representation of the i -th node in the k -th layer. The selection of f in the above formula is not arbitrary and it must be a non negative decreasing function of z^2 , in order to ensure that the loss function is non decreasing concave. Firstly, because of $\|z_i - z_j\|_2^2 = \|z_i\|_2^2 + \|z_j\|_2^2 - 2z_i^T z_j$, we can convert $f(\|z_i - z_j\|_2^2)$ to $g(z_i^T z_j)$ through variable transformation while keeping $\|z_i\|_2^2$ unchanged. In practical applications, we have added layer normalization in transformer for each layer and used a linear function $g(x) = 1 + x$. The updated function $f(\|z_i - z_j\|_2^2)$ is shown in formula(10).

$$f(\|\tilde{z}_i^{(k)} - \tilde{z}_j^{(k)}\|_2^2) = 1 + \left(\frac{\tilde{z}_i^{(k)}}{\|\tilde{z}_i^{(k)}\|_2}\right)^T \frac{\tilde{z}_j^{(k)}}{\|\tilde{z}_j^{(k)}\|_2} \quad (10)$$

We assume $\tilde{z}_i^{(k)} = \frac{z_i^{(k)}}{\|z_i^{(k)}\|_2}$, $\tilde{z}_j^{(k)} = \frac{z_j^{(k)}}{\|z_j^{(k)}\|_2}$, The above formula(10) can be written as $f(z^2) = 2 - \frac{1}{2}z^2$, this will get a non-negative result, and decreasing on the interval $[0,2]$ where z^2 is located. We bring (10) into (9) and calculate $\sum_{j=1}^N S_{ij}^{(k)} z_i^{(k)}$, as shown in formula(11):

$$\sum_{j=1}^N S_{ij}^{(k)} z_i^{(k)} = \sum_{j=1}^N \frac{1 + (\tilde{z}_i^{(k)})^T \tilde{z}_j^{(k)}}{\sum_{l=1}^N (1 + (\tilde{z}_i^{(k)})^T \tilde{z}_l^{(k)})} z_j^{(k)} = \frac{\sum_{j=1}^N z_i^{(k)} + (\sum_{j=1}^N \tilde{z}_j^{(k)} \cdot (\tilde{z}_i^{(k)})^T) \cdot \tilde{z}_i^{(k)}}{N + (\tilde{z}_i^{(k)})^T \sum_{l=1}^N \tilde{z}_l^{(k)}} \quad (11)$$

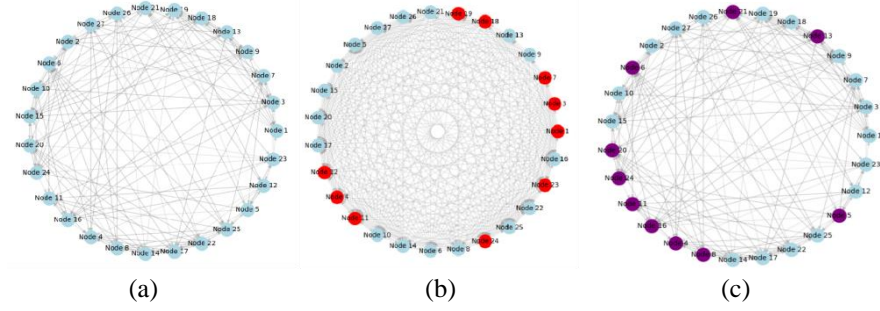


Fig.6 Avionics network topology diagram and anomaly injection

The two summation terms of the above formula can be calculated once and shared with each instance i , which can reduce the complexity of each iteration to $O(N)$. Next, we introduce the state update formula(12) to update the embedding vector of each layer node i .

$$z_i^{(k+1)} = (1 - \tau \sum_{j=1}^N \tilde{S}_{ij}^{(k)}) z_i^{(k)} + \tau \tilde{S}_{ij}^{(k)} z_j^{(k)}, 1 \leq i \leq N \quad (12)$$

Where $\tau \in (0,1)$, numerical iteration can stably converge. We can use the state after a finite number of k propagation steps for the final embedding representation. Next, we introduce the loss function as shown in formula(13).

$$L_{str}(Z, k; \sigma) = \|Z - Z(k)\|_F^2 + \lambda \sum_{i,j} \sigma(\|z_i - z_j\|_2^2) \quad (13)$$

Where σ is defined as a non-decreasing and concave function on our set interval. And enhance the robustness of the maximum difference between any pair of nodes. Formula(13) assigns an energy scalar to each state in R^d , which can be used to regularize the updated state (towards the lower energy required). The weight λ has two effects:1)

for each instance i , states not far from the current state $z_i^{(k)}$ have lower energy; 2) For all instances, the smaller the difference in their states, the lower the energy generated. Ultimately, our loss function should satisfy formula(14).

$$E(Z_{(k+1)}, k; \sigma) \leq E(Z_{(k)}, k - 1; \sigma), k \geq 1 \quad (1)$$

4 Experimental Evaluation

4.1 Dataset and outlier insertion

To our knowledge, there is currently no publicly available attribution network with true outliers. Therefore, we use the real spacecraft telemetry data of the Curiosity Mars Rover (MSL)[3] and aircraft turbine engine data(C-MAPSS), as well as injects abnormal attributes and structures[5] as our detection object, as shown in the table 1. The original dataset contains 27 sensors and 1564 temporal data. Due to the continuous nature of time series data, we can calculate the statistical characteristics of various sensor values during this period, including mean, maximum and minimum values, standard deviation, mean absolute deviation, quartile, dispersion coefficient, skewness coefficient, and kurtosis coefficient. The original 1564×27 dimensional data can be reduced to 8×27 dimensional data. This is also inline with the applicability of our proposed algorithm low dimensionality of attributes and high dimensionality of structures. Based on the Gumbel-Softmax sampling method proposed in section 3.1, construct the topology of the spacecraft sensor network, as shown in Figure 6(a).

Table 1. Dataset Introduction

Name	Nodes	Edges	Attribute dimension	outliers
MSL	27	247	9	10
C-MAPSS-T2	21	168	9	10
C-MAPSS-T67	21	189	9	8
C-MAPSS-T69	21	170	9	8
C-MAPSS-T83	21	185	9	8
C-MAPSS-T92	21	200	9	8
C-MAPSS-T96	21	195	9	8

Next, we inject random structural outliers to simulate the abnormal behavior of abnormal sensors disguised as legitimate sensor devices in the avionics system, and attempt to connect to the system for frequent interaction. As shown in Figure 6(b). Finally, we inject random attribute outliers to simulate the abnormal behavior of enemy aircraft modifying the attribute data of sensors inside our avionics system, in order to transmit

false information to the avionics system. As shown in Figure 6(c), with purple nodes representing attribute exception nodes.

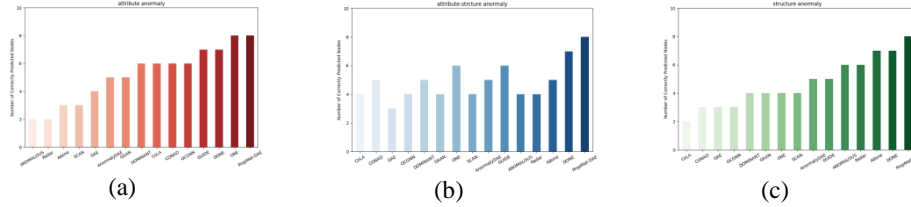


Fig.7 Baseline Algorithm Comparison Chart

4.2 Baseline

To compare and evaluate proposed algorithm, a wide range of baseline experiments were designed for each datasets. In terms of algorithm performance, we have selected 14 of the most classic attribute graph anomaly detection algorithms currently available. This includes SCAN based on clustering; Radar, ANOMALOUS, ONE based on matrix decomposition; GAE, DOMINANT, AnomalyDAE, OCGNN, CoLA, GUIDE, CONAD based on GNN principles; DONE and AdONE with MLP as the backbone; And GAAN algorithm based on GAN.

To verify the rationality of the attention mechanism designed in this article, we selected three mainstream attention mechanism design schemes, MLP, GCN, and GAT, as benchmarks. Thus, it is verified that the attention design scheme considering single layer and multi-layer neighbors is not as effective as the full attention mechanism in sensor anomaly detection tasks. To verify the rationality of the graph embedding algorithm in this article, we selected four classic graph embedding algorithms as benchmark algorithms, namely Deepwalk, Node2vec, LINE, and SDNE.

4.3 Sensor anomaly detection

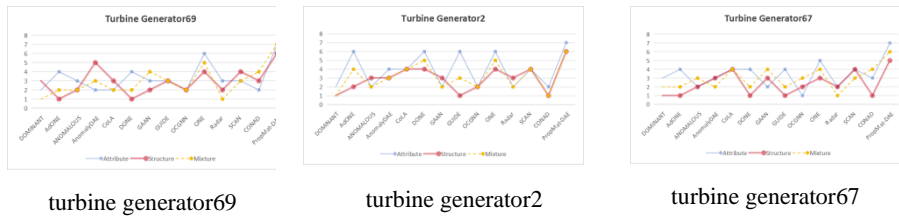
In this section, we will see the performance of all algorithms in detecting outliers we implant in the MSL dataset. PropMat-DAE directly provides the anomaly scores of each sensor node in the sensor topology diagram. Since we only injected 10 abnormal nodes, during the evaluation process, we selected the node with the highest abnormal score as the sensor fault node. By comparing the algorithm proposed in this article with the baseline graph outlier detection algorithm, the effectiveness of PropMat-DAE was verified by comparing the number of real abnormal nodes in top 10. The experimental results are shown in the figure 7. Where (a) is a node attribute. The last column in the three images is the algorithm proposed in this article. It can be observed that in attribute anomaly detection, the algorithm proposed in this article performs equally well compared to existing baseline algorithms; In structural anomaly detection, our algorithm is far ahead of other baseline algorithms; In comprehensive anomaly detection, the PropMat-DAE also produce the optimal results in sensor fault detection in avionics systems.

In addition, we also compared different datasets. NASA’s C-MAPSS (Commercial Modular Aviation Propulsion System Simulation) dataset (Turbofan Engine Degradation Simulation Dataset) is a widely used benchmark data. C-MAPSS data includes sensor data for different operating and fault conditions. Due to the inclusion of several turbine engines, this article only selects the six turbine engines (numbers are 2,67,69,83,92, and 96 respectively) with the largest sensor data for evaluation. Among them, the timing data records the sampling values of each time series during the complete cycle of the sensor from normal to fault. The comparison results of 14 algorithms are shown in the figure 8. From the graph, it can be observed that the PropMat-DAE algorithm can achieve optimal results on all six datasets, which also confirms the generalization of our algorithm in the research of sensor fault detection problems.

Due to the fact that the time series data of sensor nodes in the avionics field can be compressed into low dimensional vectors by statistical information, in order to demonstrate the rationality of our algorithm using matrix decomposition to calculate node anomaly scores for node attributes, we compared PropMat-DAE with deep learning algorithm DONE on different attribute dimensions, as shown in the figure 11. The results indicate that as the dimension of node attributes decreases, the performance of anomaly detection algorithms based on automatic encoders gradually decreases, while the matrix decomposition based gradually increases.

The PropMat-DAE is essentially an encoder. Therefore, next, we compare the performance of our algorithm with existing mainstream graph node embedding algorithms, such as deepwalk, node2vec, etc. The results are shown in the figure 9. The graph embedding algorithm only considers the graph structure without considering node features. Therefore, we can observe that in the experiment, our algorithm performs best in detecting abnormal nodes. This also proves that in the field of sensor fault detection, we should simultaneously consider the spatiotemporal attributes of nodes.

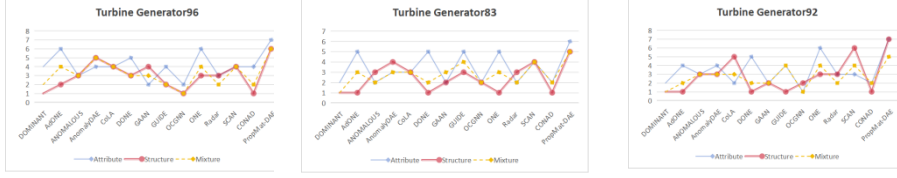
To verify the effectiveness of our designed all pair attention, we comprehensively compared the expression abilities of different attention matrices, including MLP, GCN, etc. The results are shown in the figure 10. We can observe that if we only consider the information of single-layer neighbors or their own nodes at different network layers, We tend to overlook the impact of distant nodes on target nodes. However, the interactions between various components in sensor networks are complex, so we must consider global attention. Considering the impact of all neighbors on the target node, we will achieve better results.



turbine generator69

turbine generator2

turbine generator67



turbine generator96

turbine generator83

turbine generator92

Fig.8. C-MAPSS data

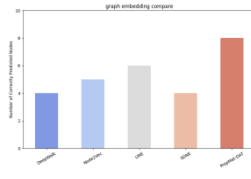


Fig.9 embeddings

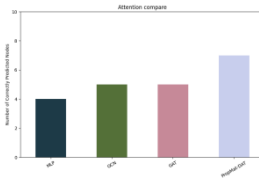


Fig.10 attentions

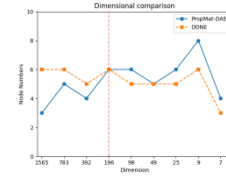


Fig.11 dimensions

5 Conclusion

In this article, we introduce PropMat-DAE, a sensor fault detection framework that improves on previous graph anomaly detection algorithms. After optimization, we use matrix decomposition to calculate the reconstruction error of low dimensional attribute information, and use the all pair attention mechanism combined with AE thinking to calculate the high-dimensional reconstruction error of the manuscript, thereby calculating the node anomaly score. In the experiment, we validated its effectiveness over the latest graph anomaly detection algorithms and graph embedding algorithms. It also proves the feasibility of matrix decomposition and all pair attention. Although our proposed algorithm has achieved good results in this article, our goal in the future is to provide more rigorous analysis and further achieve sensor fault detection across different fields.

However, this article still has some limitations. Firstly, the dataset selected in this article is relatively small and not suitable for large-scale datasets. Secondly, the graph dataset used in this article was calculated by the algorithm and may differ significantly from the actual structure. Therefore, in terms of engineering implementation, it is important to focus on the actual topology structure. Finally, the global attention mechanism designed in this article has a relatively large time consumption, but performs well in information aggregation. Therefore, the all air attention in this article is not suitable for large-scale graph data.

Finally, the contribution made in this article is mainly reflected in considering the abnormal situation of sensor nodes from both time and space perspectives, which compensates for the shortcomings of most sensor anomaly detection. Moreover, due to the

high confidentiality of spacecraft topology, the graph structure learning algorithm introduced in this paper can also provide new ideas for this limitation and lay the foundation for subsequent related research.

References

1. Chen, S., Yang, R., Zhong, M., Xi, X., Liu, C.: A random forest and model-based hybrid method of fault diagnosis for satellite attitude control systems. *IEEE Transactions on Instrumentation and Measurement* 72, 1–13 (2023). <https://doi.org/10.1109/TIM.2023.3279453>
2. Hasan, A., Tahavori, M., Midtby, H.S.: Model-based fault diagnosis algorithms for robotic systems. *IEEE Access* 11, 2250–2258 (2023). <https://doi.org/10.1109/ACCESS.2022.3233672>
3. Hundman, K., Constantinou, V., Laporte, C., Colwell, I., Soderstrom, T.: Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM (jul 2018). <https://doi.org/10.1145/3219819.3219845>
4. Li, T., Sun, C., Li, S., Wang, Z., Chen, X., Yan, R.: Explainable graph wavelet denoising network for intelligent fault diagnosis. *IEEE Transactions on Neural Networks and Learning Systems* (2022)
5. Liu, K., Dou, Y., Zhao, Y., Ding, X., Hu, X., Zhang, R., Ding, K., Chen, C., Peng, H., Shu, K., Sun, L., Li, J., Chen, G.H., Jia, Z., Yu, P.S.: Bond: Benchmarking unsupervised outlier node detection on static attributed graphs (2022)
6. Maddison, C.J., Tarlow, D., Minka, T.: A* sampling (2015)
7. Rao, M., Zuo, M.J., Tian, Z.: A speed normalized autoencoder for rotating machinery fault detection under varying speed conditions. *Mechanical Systems and Signal Processing* 189, 110109 (2023). <https://doi.org/https://doi.org/10.1016/j.ymsp.2023.110109>
8. Sun, Z., Wang, Y., Sun, G., Gao, J., Pan, Z.: Neighborhood graph embedding interpretable fault diagnosis network based on local and non-local information balanced under imbalanced samples. *Structural Health Monitoring* 22(3), 1721–1744 (2023)
9. Yadegar, M., Bakhtiaridoust, M., Meskin, N.: Adaptive data-driven fault-tolerant control for nonlinear systems: Koopman-based virtual actuator approach. *Journal of the Franklin Institute* 360(11), 7128–7147(2023). <https://doi.org/https://doi.org/10.1016/j.jfranklin.2023.05.029>
10. Bangyal, W.H.; Nisar, K.; Ag, Ibrahim, A.A.B.; Haque, M.R.; Rodrigues, J.J.P.C.; Rawat, D.B. Comparative Analysis of Low Discrepancy Sequence-Based Initialization Approaches Using Population-Based Algorithms for Solving the Global Optimization Problems. *Appl. Sci.* 2021, 11, 7591. <https://doi.org/10.3390/app11167591>
11. A. Anwaar, A. Ashraf, W. H. K. Bangyal and M. Iqbal, "Genetic Algorithms: Brief review on Genetic Algorithms for Global Optimization Problems," 2022 Human-Centered Cognitive Systems (HCCS), Shanghai, China, 2022, pp. 1-6, doi: 10.1109/HCCS55241.2022.10090327.
12. W. H. Bangyal, Z. A. Malik, I. Saleem and N. U. Rehman, "An Analysis of Initialization Techniques of Particle Swarm Optimization Algorithm for Global Optimization," 2021 International Conference on Innovative Computing (ICIC), Lahore, Pakistan, 2021, pp. 1-7, doi: 10.1109/ICIC53490.2021.9692931.

