Asia Pacific Science Press

Asia Pacific Economic and Management Review
ISSN(O):3005-9275
ISSN(P):3005-9267
Vol. 3 No.1 (2026)

Anomaly Detection of Collusion Bidding in Electricity Market Based on Deep Learning Model

Ziyu Zhao*

International Business School, Hainan University, Haikou, Hainan 570228, China

Corresponding author: Ziyu Zhao

Abstract: In the electricity market, collusion bidding involves multiple entities colluding to inflate prices or conceal capacity, which disrupts marginal cost pricing. Traditional rules and single-variable thresholds struggle to detect such collusion in a timely manner. This paper addresses a real-world scenario with only six input features and highly imbalanced labels by constructing an anomaly detection framework that combines Transformer-Autoencoder (TAE) with Gaussian Mixture Model (GMM). The model uses multi-head self-attention to capture the coupling relationships among original input features, measures sample rarity using Gaussian mixture density in the space of latent variables and reconstruction errors, and automatically generates an energy threshold at the 90th percentile of the normal distribution for anomaly detection. Experimental results show that on the electricity collusion dataset, this method achieves detection performance with Precision 0.80, Recall 0.79, F10.79, and AUC 0.844. It is only ±2% sensitive to fluctuations in the energy weight λ, demonstrating robust performance. The energy distribution compresses normal sample data into spikes while pushing anomalies to the long tail, achieving efficient anomaly detection. Attention heat maps and gradient sensitivity both point to feature one as the most critical feature, validating the interpretability of the model's decision logic. TAE-GMM can be trained solely using normal samples, and the final detection threshold is determined by the proportion of normal samples, enabling simple and flexible application.

Keywords: Electricity Collusion; Anomaly Detection; GMM; Transformer

Published: Feb 24, 2026

DOI: https://doi.org/10.62177/apemr.v3i1.1068

1.Introduction

1.1 Research Background and Importance

In recent years, the global power industry has been accelerating its transition from a vertically integrated, utility-dominated planned system to a market-oriented generation bidding mode [1]. In this context, as a special commodity with instantaneous production-consumption and difficulty in large-scale storage, electricity should theoretically rely on full competition to achieve marginal cost pricing, thereby improving resource allocation efficiency and reducing end-user electricity prices [2]. However, experience and cases have shown that "deregulation" does not truly equate to "efficiency": when generation-side capacity is concentrated and demand-side price elasticity is limited, individual entities have the motivation and ability to manipulate the System Marginal Price (SMP/LMP) [3]. A more covert practice is that multiple generating units collude to raise quotes or conceal available capacity during critical load periods, also known as collusion behavior, whose characteristics often present as “normal single-point indicators but abnormal combined features”, making traditional monitoring methods based on thresholds or univariate statistics ineffective [4].

In the past, regulatory authorities mainly relied on expert experience rules or post-event investigations to identify such collaborative behaviors. However, as manipulation strategies continue to evolve, this approach has shown problems of lag and limited coverage [5]. With the continuous digitalization of metering devices, dispatching automation, and trading systems, the electricity market has accumulated massive high-frequency bidding records, laying a foundation for data-driven real-time anomaly monitoring [6]. Deep Learning (DL), especially autoencoders and generative models, has demonstrated strong insight into low-probability anomalies in unsupervised or weakly supervised scenarios [7]. Nevertheless, traditional convolutional or simple fully connected networks still struggle in electricity bidding scenarios with “low feature dimensions and complex inter-sample coupling” [8].

The Transformer structure can automatically capture high-order dependencies through the multi-head self-attention mechanism, having achieved cross-domain generalization effects in fields such as language, speech, and industrial time-series signals [9]-[10], which provides a new idea for electricity market anomaly detection. Introducing Transformer into inputs containing only six core bidding features can resolve potential coupling relationships without additional prior knowledge, thereby identifying hidden collusion patterns [11].

Therefore, targeting the electricity market collusion dataset, this paper constructs a TAE-GMM framework integrating Transformer self-attention and Gaussian mixture density estimation. It systematically evaluates the contribution of each feature in anomaly detection through multi-dimensional interpretation methods including gradient sensitivity and attention mapping. This paper not only responds to the practical demand of the electricity market for efficient, interpretable, and scalable monitoring tools, but also demonstrates the application potential of generative deep models in small-feature representation scenarios for other safety-critical industries.

2.Algorithm Principles

The algorithm framework consists of four sequentially connected subsystems: Transformer encoder, decoder, energy estimation network, and Gaussian Mixture Model. The latent vector compressed by the Transformer-Autoencoder $\mathbf{z}$ records the comprehensive pattern of the sample in the original feature space, while the additional reconstruction error $e = \| \mathbf{x} - \hat{\mathbf{x}} \|_2^2$ captures the difficulty encountered by the network in reconstructing the sample. After concatenating the two into $\mathbf{v} = [\mathbf{z}; e]$, if the vector falls within the high-probability region of “normal behavior” learned during training, it is considered normal; once it is located in the density tail, it is regarded as an outlier.

First, the original feature vector of length 6 $\mathbf{x}$ is split into F+1=7 tokens after column embedding and scalar projection, where the 0th position is [CLS]:

$$\mathbf{T} = \left[\mathbf{e}_{[\mathrm{CLS}]} \,\|\, W_s x_1, \ldots, W_s x_6\right] + \mathbf{E}_{\mathrm{col}} \tag{2,1}$$

where $W_s \in \mathbb{R}^{1 \times d_{\mathrm{model}}}$ is the shared scalar projection matrix and $\mathbf{E}_{\mathrm{col}} \in \mathbb{R}^{7 \times d\mathrm{model}}$ is the learnable column embedding. Subsequently, T passes through L layers of Multi-Head Self-Attention+Feed-Forward Network (MHSA+FFN) modules in sequence to capture inter-column collaboration. The Transformer sub-structure can be abbreviated as:

$$\mathbf{H}^{(l)} = \mathrm{FFN}^{(l)}\left(\mathbf{H}^{(l-1)} + \mathrm{MHSA}^{(l)}\left(\mathbf{H}^{(l-1)}\right)\right), \quad l = 1, \ldots, L, \tag{2,2}$$

Finally, the normalized output $\mathbf{H}^{(L)}$ s obtained, and the [CLS] token is extracted as the global semantic vector $\mathbf{h}_{\mathrm{CLS}}$. The latent variable is obtained through linear projection dimensionality reduction:

$$\mathbf{z} = W_p \mathbf{h}_{\mathrm{CLS}} + \mathbf{b}_p \quad W_p \in \mathbb{R}^{d_{\mathrm{model}} \times d_z} \tag{2,3}$$

The decoder is a symmetric Multi-Layer Perceptron (MLP) used to reconstruct the input and generate reconstruction error:

$$\begin{aligned} \hat{\mathbf{x}} &= g_\phi(\mathbf{z}) \\ e &= \| \mathbf{x} - \hat{\mathbf{x}} \|_2^2 \end{aligned} \tag{2,4}$$

To highlight the importance of the error in the discrimination stage, e and z are concatenated to form the density estimation space vector:

$$\mathbf{v} = [\mathbf{z}\,;\, e] \in \mathbb{R}^{d_z + 1} \tag{2,5}$$

The estimation network ψ_η contains only one tanh layer followed by SoftMax, and its output $\boldsymbol{\gamma} = \psi_\eta(\mathbf{v})$ is the responsibility of each sample belonging to the k-th mixture component. The iterative update formulas for Gaussian Mixture Model parameters are as follows:

$$\begin{aligned}
\gamma_{ik} &= \frac{\pi_k \mathcal{N}\left(\mathbf{v}_i \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)}{\sum_{j=1}^{K} \pi_j \mathcal{N}\left(\mathbf{v}_i \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\right)} \\
\pi_k &= \frac{1}{N} \sum_{i=1}^{N} \gamma_{ik} \\
\boldsymbol{\mu}_k &= \frac{1}{N_k} \sum_{i=1}^{N} \gamma_{ik} \mathbf{v}_i \\
\boldsymbol{\Sigma}_k &= \frac{1}{N_k} \sum_{i=1}^{N} \gamma_{ik} \left(\mathbf{v}_i - \boldsymbol{\mu}_k\right)\left(\mathbf{v}_i - \boldsymbol{\mu}_k\right)^{\top}
\end{aligned} \tag{2,6}$$

The resulting energy score is defined as:

$$\mathcal{E}(\mathbf{v}) = -\ln\Big[\sum_{k=1}^{K} \pi_k \mathcal{N}\left(\mathbf{v} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)\Big] \tag{2,7}$$

A higher energy indicates that the sample is rarer in the latent distribution, thus serving as an anomaly metric.
During the training phase, the total loss function is a combination of reconstruction loss and energy:

$$\mathcal{L} = \underbrace{\frac{1}{N} \sum_{i=1}^{N} \| \mathbf{x}_i - \hat{\mathbf{x}}_i \|_2^2}_{\text{Reconstruction Loss}} + \lambda \underbrace{\frac{1}{N} \sum_{i=1}^{N} \mathcal{E}\left(\mathbf{v}_i\right)}_{\text{Energy}} \tag{2,8}$$

By controlling the hyperparameter λ, the overall training objective balances both reconstruction and density aspects. The encoder is encouraged to accurately reconstruct while being driven to compress normal samples into the "core region" of multi-modal density, thereby making the energy score highly sensitive to anomalies.
To map continuous energy to "anomaly/normal" labels, we set the 90th percentile of the energy distribution of the validation set containing only normal samples as the threshold τ, since the proportion of normal values in the original dataset is approximately 90%. In actual prediction, this percentile can be modified. Since the threshold is completely determined by normal data, the model remains sensitive to unseen anomaly patterns in the future. Meanwhile, using a single real number τ allows business parties to easily adjust the trade-off between false negatives and false positives. With the density estimation of GMM, we map high-dimensional deep features to a physically interpretable probability space. Then, through the energy score and percentile-based threshold strategy, robust and flexible online anomaly detection is achieved.

3.Research Results

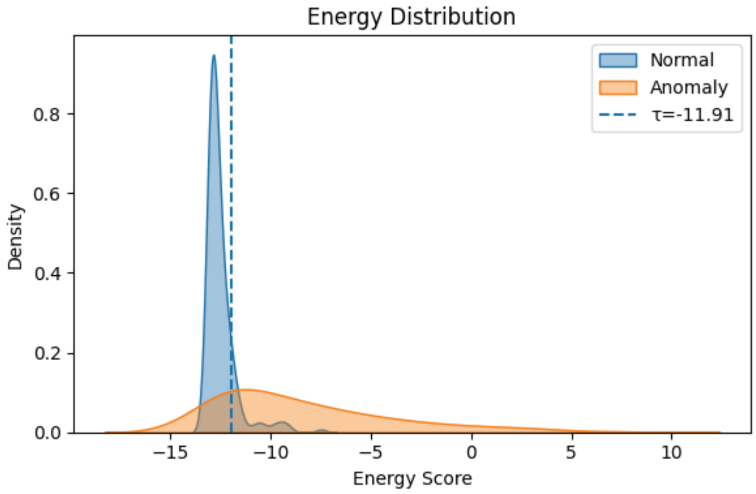
To explore the impact of different values of λ on the final experimental results, repeated experiments were conducted with values of λranging from 0.005 to 0.1, and the results are shown in Table 1. It can be seen that the best performance is achieved when λ= 0.05, with an AUC value of 0.844.

Table 1 Experimental results with different values of λ

λ	**AUC**	**Precision**	**Recall**	**F1**
0.005	0.835093	0.792453	0.770642	0.781395
0.01	0.830506	0.790476	0.761468	0.775701
0.05	0.844267	0.796296	0.788991	0.792627
0.1	0.839680	0.794393	0.779817	0.787037

For visual analysis, the energy score distribution diagram was plotted, as shown in Figure 1. The energy scores obtained during the validation-test phase are divided into two kernel density curves: the blue curve represents samples judged as normal by the model (operating modes seen during training), and the orange curve represents known collusion samples. The threshold τ≈−11.9 shown by the dashed line is the 90th percentile taken on the validation set containing only normal samples. That is, only 10% of normal behaviors fall to the right of the dashed line. In actual monitoring, all records with energy exceeding τ will be marked as anomalies.

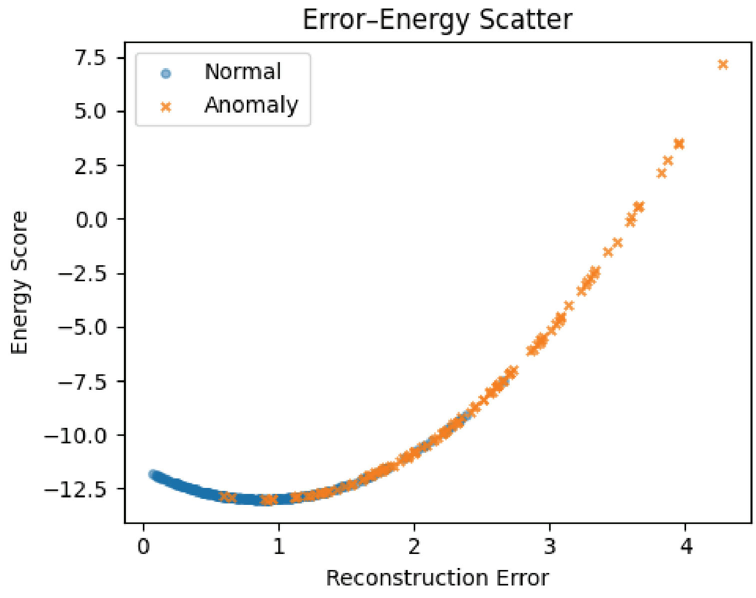
Figure1 Energy density distribution



The steep blue curve indicates that normal samples are very compact in the latent density space. In contrast, the orange curve "tails" to the right and crosses the dashed line, indicating that collusion bidding is significantly sparse in the energy dimension, with a large number of samples falling into the "alarm zone" to the right of the threshold. There is also a small overlap of the orange distribution on the left shoulder of the blue peak, meaning that a few anomaly points still have energy values close to those of normal samples, which are difficult to distinguish and represent detection challenges for the model.

The reconstruction error and energy of each test record are projected onto a two-dimensional plane to draw a scatter plot, as shown in Figure 2. The horizontal axis is the reconstruction error of the autoencoder (the larger the value, the more difficult it is for the network to restore the input to its original form), and the vertical axis is the energy score calculated by GMM (the larger the value, the rarer the sample in the normal density).

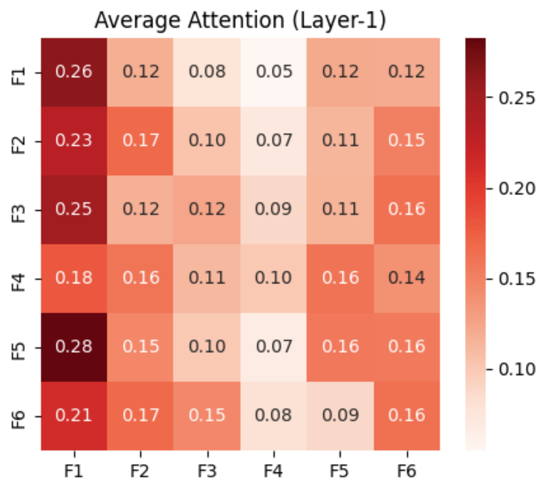
Figure2 Reconstruction loss and energy



Most normal points are concentrated in the lower-left valley, while anomaly points spread along a parabola to the upper right. The further away from the valley, the higher the probability of anomaly. This indicates that the selection of λ at this time is reasonable, which can distinguish most normal samples from abnormal data.

To further explore the importance of different features in actual anomaly detection, a heatmap of the average attention intensity of the first layer of the Transformer encoder for the dependency relationships between the six original features was plotted. The horizontal axis represents the attended (Key) features, and the vertical axis represents the attending (Query) features. The darker the color and the larger the value, the more the encoder tends to rely on the information of the corresponding column when encoding the Query feature.

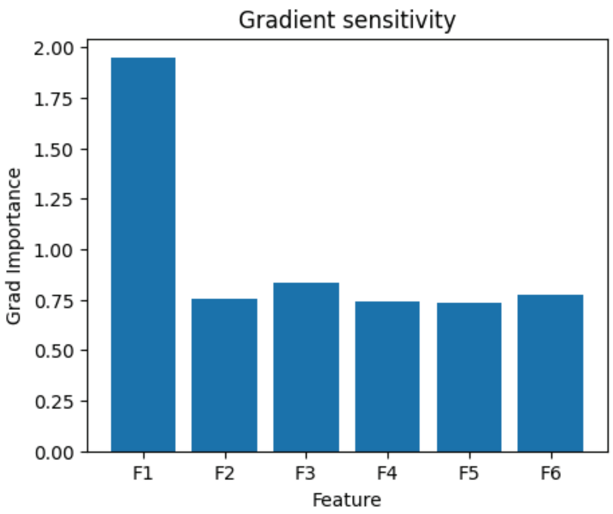
Figure3 Average attention scores



The darkest color band in Figure 3 is located in the first column: F2, F3, F5, and F6 all give attention to F1 ranging from 0.21 to 0.28. This indicates that when encoding other features, the model primarily refers to the information of Feature 1, which is a feature that needs to be mainly focused on in subsequent practical applications. Among them, the attention intensity of F4 (i.e., Feature 4) is significantly lower than that of other features, indicating that the data of this feature contributes little to distinguishing outliers.

For each test sample, this paper calculated the partial derivatives of the energy score $\mathcal{E}(\mathbf{x})$ with respect to the six original features, and then took the average of the absolute values across all samples. The larger the value on the vertical axis, the more significant the impact of minor perturbations of the feature on the energy change in the model's perspective, and thus the more critical the feature, as shown in Figure 4.

Figure4 Gradient sensitivity



The results show that the height of the bar for the first column F1 is 1.95, which is significantly higher than the other five columns of features. This means that a slight modification of F1 will cause a significant shift in the sample's position in the GMM density, and the anomaly confidence will fluctuate greatly. The results are consistent with the aforementioned attention scores. From the perspective of anomaly detection, the gradient sensitivity diagram provides a clear priority ranking: to quickly screen out most potential collusion, first closely monitor the value of the first feature.

4.Discussion

Comprehensive quantitative indicators and visualization results show that the Transformer-Autoencoder+GMM (TAE-GMM) framework proposed in this paper not only meets the accuracy requirements of market regulation in terms of pure detection performance, but also provides a clear and traceable evidence chain at the interpretation level. Firstly, the performance comparison of different energy weights λ in Table 1 reveals the optimal balance between reconstruction and density regularization. When $\lambda=0.05$, the model achieves Precision 0.844, Recall 0.789, F1 0.793, and AUC 0.796 simultaneously. Further increasing or decreasing λ will sacrifice the performance of one end. Notably, the indicator fluctuations in the four groups of experiments are only within ±2%, indicating that the framework is not sensitive to hyperparameters. When migrating to other regions or larger samples, it can be put into use with only lightweight parameter tuning. The energy distribution diagram compresses normal samples into sharp and narrow density peaks, while collusion samples form a long tail and significantly cross the threshold τ. This confirms the modeling capability of GMM, which can steadily wrap normal clusters while throwing abnormal collusion samples to the tail. The Error-Energy scatter plot further illustrates that reconstruction error and energy score are not redundant signals. Some anomaly samples are not prominent in one dimension, but are significantly far from the normal valley in the two-dimensional plane, proving the complementarity of the two indicators.

5.Conclusion

This paper uses the self-attention mechanism of Transformer to characterize the nonlinear coupling between different features, and then uses GMM energy score to measure anomaly rarity. It achieves high-precision prediction with F1=0.79 on the electricity collusion dataset, and verifies the interpretability of the detection basis through a series of visualizations. The model only needs normal data for training, and the corresponding threshold τ can be directly generated by adjusting the proportion of normal data and abnormal data, thereby realizing flexible anomaly sample detection without strictly relying on a small number of abnormal samples.

Funding

No

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

[1] European Parliament Think Tank. (2023). Improving the design of the EU electricity market.

[2] Twomey, P., et al. (2005). A review of the monitoring of market power. MIT Center for Energy and Environmental Policy Research.

[3] Brown, D., Eckert, A., & Silveira, D. (2023). Screening for collusion in wholesale electricity markets: A review.

[4] Synapse Energy Economics. (2007). LMP electricity markets: Market power case studies.

[5] MIT Center for Energy and Environmental Policy Research. (2023). Calls for an electricity market reform in the EU.

[6] Frontiers in Energy Research. (2021). Power consumption predicting and anomaly detection based on transformer and K-means.

[7] PeerJ Computer Science. (2023). Detecting anomalous electricity consumption with transformer and synthesized anomalies.

[8] Wang, S., et al. (2025). Self-attention-based graph transformation learning for anomaly detection in multivariate time series.

[9] Mu, T., et al. (2023). Transformer-based anomaly detection for electricity theft. PeerJ Computer Science.

[10] Zhang, X., et al. (2025). Domain adaptation of attention heads for zero-shot anomaly detection. arXiv.

[11] Zhang, J., et al. (2021). Power consumption predicting and anomaly detection based on transformer and K-means. Frontiers in Energy Research.