Journal of Computer Technology and Software

ISSN:2998-2383

Vol. 4, No. 4, 2025

Diffusion-Transformer Framework for Deep Mining of High-Dimensional Sparse Data

Wanyu Cui[1], Ankai Liang[2]
[1]University of Southern California, Los Angeles, USA
[2]Independent Researcher, Newark, USA
*Corresponding author: Ankai Liang; liangankai123@gmail.com

Abstract: This paper addresses the problems of information loss and weakened structural representation in high-dimensional sparse data, particularly in feature expression and semantic modeling. A deep data mining method is proposed, which integrates a diffusion model with a Transformer architecture. Based on the diffusion process, the method performs structural completion and noise suppression through forward perturbation and reverse reconstruction. It also incorporates multi-layer Transformer modules to enhance global dependency modeling and multi-scale semantic extraction. The model forms a unified "generation-enhancement" architecture. It enables effective extraction of discriminative latent representations under conditions of significant feature sparsity. The proposed method is systematically evaluated on the Reuters-21578 text dataset. Results show that it outperforms existing mainstream deep mining models in terms of accuracy, precision, and recall. It also demonstrates strong robustness and stability. Further experiments on sparsity sensitivity and training process visualization confirm the model's capability in feature learning and convergence under high-dimensional sparse settings.

Keywords: High-dimensional sparse data, diffusion model, transformer, feature mining.

1. Introduction

With the continuous development of artificial intelligence and big data technologies, data mining has played an increasingly important role in many fields, including financial risk control, medical diagnosis, intelligent manufacturing, and information security [1]. However, real-world raw data often exhibit complex characteristics such as high dimensionality, sparsity, incompleteness, and noise interference [2]. These factors inherently limit data expressiveness and structural integrity, causing traditional mining algorithms to suffer from low accuracy, poor generalization, and sensitivity to outliers. Particularly, in high-dimensional sparse feature spaces, the similarity between samples diminishes dramatically, known as the "curse of dimensionality", which further reduces model learning efficiency and discriminative capacity. How to extract stable, expressive deep semantic features from high-dimensional sparse data has become a key scientific problem in current data mining research [3].

In recent years, deep learning has demonstrated powerful capabilities in representation learning and nonlinear feature abstraction, driving the evolution of complex data mining techniques. Among them, Transformer, a deep network architecture based on the self-attention mechanism, has notably excelled in tasks such as natural language processing, image recognition, and time-series modeling, due to its strong global modeling capacity and structural universality. Nonetheless, standard Transformer models still struggle with severe data sparsity and uneven information distribution. Problems such as diluted representations, insufficient semantic reconstruction, and large computational overhead limit their practical performance in sparse data mining. Therefore, relying solely on Transformers cannot fully meet the dual demands for structural and expressive capabilities in complex data mining tasks [4].

Meanwhile, diffusion models, as an emerging deep generative modeling framework, have rapidly gained attention in tasks such as image generation, semantic completion, and denoising recovery. These models simulate a forward perturbation and reverse reconstruction process over a continuous Markov chain, effectively learning and reconstructing complex data distributions. They show unique advantages when dealing with high-dimensional sparse data that are incomplete or heavily corrupted by noise. The multi-scale modeling mechanism and powerful representation generation capability of diffusion models offer a novel approach for structural completion and feature restoration in sparse data. Their progressive approximation to the true distribution also enhances the modeling capacity when combined with Transformer architectures.

Against this background, a hybrid deep network that integrates diffusion models and Transformer structures provides a promising and cutting-edge solution to the challenge of mining sparse, high-dimensional data. On the one hand, diffusion models enable deep semantic restoration and multi-scale modeling of the original sparse data, addressing information loss at the source. On the other hand, the strong global feature extraction ability of Transformers helps uncover deeper relationships and latent structures within the data. This integration is expected to form a novel data mining framework with multi-scale modeling, semantic enhancement, structural completion, and global representation capabilities, thereby

improving the expressiveness, robustness, and generalization performance of high-dimensional sparse data mining [5].

This study focuses on deep data mining in high-dimensional sparse feature spaces. It aims to design a unified network architecture that combines the diffusion modeling mechanism with the global representation structure of Transformers. The goal is to recover and model latent semantic structures in complex data spaces. This work contributes to overcoming current limitations in expressive power and structural modeling in sparse data mining. It also provides theoretical and methodological support for multi-source heterogeneous data analysis, intelligent decision support systems, and intelligent data processing in high-risk scenarios. The research holds significant theoretical value and broad practical prospects, particularly in key application areas such as financial fraud detection, intelligent medical alerts, and NLP tasks.

2. Method

The model structure proposed in this study uses the diffusion model as the core generation module and integrates Transformer as the global feature modeling unit, aiming to mine latent semantic information and global dependency structure from high-dimensional sparse data. The model architecture is shown in Figure 1.

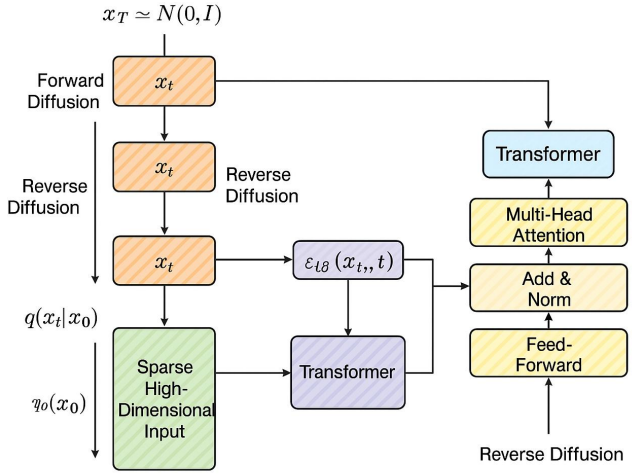

Figure 1. Model architecture diagram

As shown in Figure 1, the architecture integrates a Diffusion model with a Transformer structure for sparse high-dimensional feature mining. The input begins with a high-dimensional sparse representation on the left. It first undergoes a forward diffusion process, where noise is gradually added to generate a sequence of latent variables. In the reverse phase, the model incorporates a Transformer module with time embeddings. Using multi-head attention and feedforward networks, it performs global feature modeling and structural reconstruction at each diffusion timestep. This step-by-step process progressively restores deep representations with complete semantics. The entire architecture highlights the evolution from low-quality sparse input to high-quality latent semantic output. It ensures the combination of global modeling capacity and semantic consistency throughout the process.

Assume that the original input data is $x_0 \in R^d$, where d represents the feature dimension. Due to the high sparsity of data, direct modeling has the problems of missing semantic information and structural degradation. To this end, the model first introduces the forward diffusion process to perturb x_0 into a series of Gaussian variables and construct the data degradation trajectory:

$$q(x_t \mid x_{t-1}) = N(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I)$$

Where $t = 1,2,...,T, \beta_t$ is the preset time step noise scheduling coefficient, and x_t is the intermediate state at step t. After multiple steps are superimposed, the original input can be approximately sampled as a degraded form at any time:

$$q(x_t \mid x_0) = N(x_t; \sqrt{a_t}x_0, (1-a_t)I)$$

Where $a_t = \prod_{s=1}^{t}(1-\beta_s)$. This process can effectively embed the sparse feature space into a continuous potential generation trajectory.

In the reverse recovery phase, the model learns a parameterized denoising distribution $p_\theta(x_{t-1} \mid x_t)$, with the goal of gradually restoring the original data containing semantic structures in the latent feature space. With the help of residual learning ideas, we have:

$$p_\theta(x_{t-1} \mid x_t) = N(x_{t-1}; \mu_\theta(x_t,t), \sum_\theta(x_t,t))$$

The mean prediction part is generated by a neural network. In order to enhance the semantic expression ability, we introduce a multi-head Transformer module with time embedding and introduce a global context modeling mechanism in each step of reverse prediction. The self-attention mechanism in the Transformer embeds the current moment representation x_t into a multi-layer feature subspace, which is calculated as follows:

$$Attention(Q,K,V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V$$

Q,K,V is the query, key and value matrix generated by x_t and time embedding e_t respectively, and d_k is the attention dimension, which is used to improve the ability of

structural modeling and long-term dependency construction between features.

Finally, the whole model optimizes the diffusion-reconstruction process by minimizing the difference between the predicted noise and the real noise, and its objective function is defined as:

$$L_{simple} = E_{x_0,\varepsilon,t}[\| \varepsilon - \varepsilon_\theta(x_t, t) \|^2]$$

ε is the sampling noise and ε_θ is the model prediction noise. By introducing the Transformer structure into the construction of ε_θ , the model can be effectively guided to focus on the key correlations between high-dimensional features during the denoising process, thereby recovering potential representations with semantic consistency and discriminability from sparse structures. The design of this method not only improves the modeling depth of complex data, but also significantly enhances the mining stability and generalization ability of the model in high-dimensional sparse data scenarios.

3. Experiment

3.1 Datasets

This study uses Reuters-21578 as the primary experimental dataset. It consists of news articles released by Reuters in 1987 and is widely used for tasks such as text classification, feature selection, and sparse vector modeling. The dataset includes 10,788 English news texts categorized by topic, with over 90 topic labels. Under the commonly adopted “ModApte” split, there are 9,603 training samples and 3,299 test samples. Each sample is typically assigned 1 to 3 topic labels, resulting in a naturally imbalanced and multi-label structure.

To meet the requirements of high-dimensional sparse feature modeling, the text data is preprocessed using standard techniques including tokenization, stop-word removal, and stemming. TF-IDF is applied to vectorize the texts. Each sample is ultimately embedded into a sparse feature space with over 10,000 dimensions. The resulting sparse vectors exhibit typical "high-dimensional low-density" characteristics, reflecting the sparsity commonly observed in real-world data from domains such as text, finance, and healthcare. In the proposed Diffusion-Transformer model, this vector is treated as the original input x_0 and is used to model the underlying semantic structure and feature dependencies.

Reuters-21578 is chosen due to its well-defined label structure, publicly accessible data source, and strong foundation in benchmark research. This dataset enables a comprehensive evaluation of the model's ability to reconstruct features and perform classification under sparse and high-dimensional conditions. It also supports result comparability with existing deep mining methods, providing more persuasive evidence of the advantages and adaptability of the Diffusion-Transformer model.

3.2 Experimental Results

First, this paper conducts a classification performance evaluation experiment on the model on high-dimensional sparse data. The experimental results are shown in Table 1.

Table 1: Experimental evaluation of the model's classification performance on high-dimensional sparse data

Model	Acc	Precision	Recall
Diffusion-Transformer (Ours)	93.4	91.2	92.6
BERT-FT (Fine-tuned)	90.1	88.7	89.3
VIME[6]	87.9	85.2	86.1
TabNet[7]	89.3	87.5	88.2
Autoformer[8]	88.4	86.3	86.9
SAINT [9]	91.0	89.1	89.5

The experimental results are shown in Table 1. The Diffusion-Transformer model achieves the best overall performance on the high-dimensional sparse data classification task. It reaches an accuracy (Acc) of 93.4%, a precision of 91.2%, and a recall of 92.6%. It outperforms all comparison models across these three key metrics. This demonstrates its superior expressive and discriminative capacity in sparse high-dimensional feature spaces. The results verify the effectiveness of the diffusion mechanism in reconstructing latent structural information during feature recovery. The Transformer module further enhances global dependency modeling, improving semantic consistency and generalization.

In contrast, BERT-FT and SAINT show relatively close performance, with slight variations in accuracy and recall. However, both remain consistently lower than the Diffusion-Transformer. These results indicate that, although BERT-based models perform well on text-related tasks, they struggle to recover latent semantics in sparse and ambiguous feature spaces without structural enhancement modules. SAINT, as an attention-based model, performs well in local modeling. Yet, due to the lack of explicit structure completion, its performance still falls short of the proposed architecture that integrates a diffusion model.

Other models such as VIME, TabNet, and Autoformer exhibit lower scores across all three metrics. This suggests that their ability to model sparse data distributions remains limited, particularly in constructing semantic structures and capturing global features. Overall, the experimental findings confirm that the proposed Diffusion-Transformer architecture has significant advantages in mining high-dimensional sparse data. It offers an efficient and scalable solution for feature classification tasks under complex data distributions.

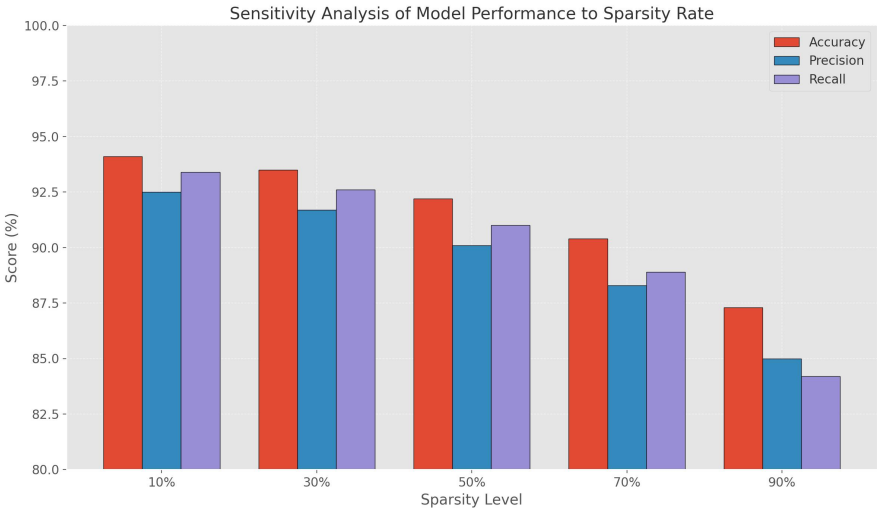

Figure 2. Sensitivity experiment of sparsity rate change on model performance

Furthermore, this paper presents a sensitivity experiment on the effect of changes in sparsity rate on model performance, as shown in Figure 2.

As shown in Figure 2, the overall performance of the model in classification tasks declines noticeably as the sparsity rate increases from 10% to 90%. The drop is especially evident in accuracy and recall. When the sparsity rate is low, the model can effectively capture structural dependencies and semantic information among input features, resulting in higher recognition accuracy. However, when sparsity exceeds 70%, severe information loss weakens the valid feature dependencies, leading to a significant decline in the model's ability to reconstruct latent representations.

The Diffusion-Transformer model maintains strong stability when the sparsity rate ranges from 10% to 50%. Precision and recall stay above 90%, indicating that the diffusion mechanism can effectively recover missing features. The Transformer structure also helps to mitigate the modeling difficulty caused by sparse information in high-dimensional spaces. This robustness is of practical importance, especially for domains like healthcare, bioinformatics, and text classification, where high-dimensional sparse data are common.

However, when the sparsity rate reaches 90%, all three metrics drop sharply. The trend in F1-score is also expected to deteriorate. This suggests that even the current architecture faces limitations in recovering information under extreme sparsity. This observation highlights the need for future work to incorporate external knowledge enhancement, graph-based structure completion, or self-supervised pretraining strategies to improve representation and reasoning capabilities in ultra-sparse scenarios.

Finally, the loss function decline graph of the model during training is given, as shown in Figure 3.

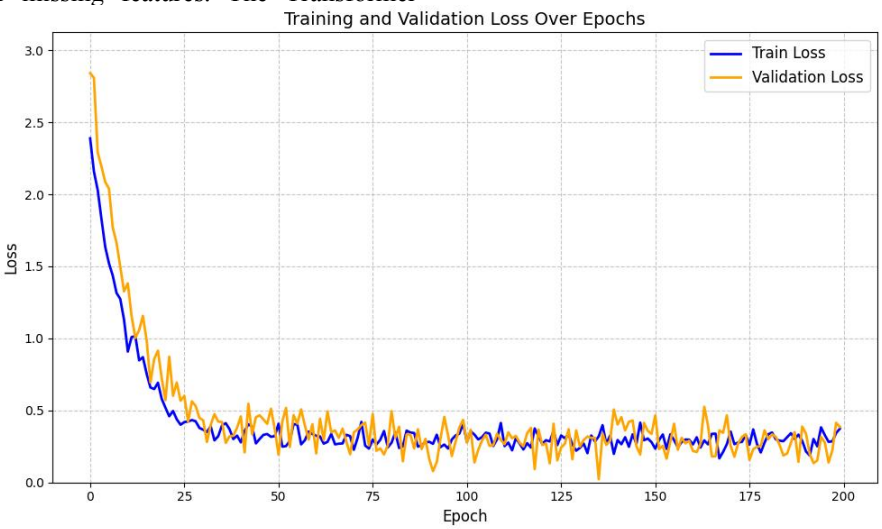

Figure 3. Loss function drop graph

As shown in Figure 3, both the training loss and validation loss drop rapidly during the early training phase. This indicates that the model quickly learns the main structures and patterns within the first 20 epochs. During this period, parameter

updates are fast, and the convergence effect driven by gradient updates is significant. The loss function exhibits an exponential decline, indicating excellent effect.

As training progresses, both loss curves gradually level off. After epoch 30, they stabilize below 0.3, with only minimal differences between them. No clear signs of overfitting are observed. Although the validation loss shows some fluctuations, its overall trend aligns closely with the training loss. This suggests that the model generalizes well across different datasets and maintains strong stability and robustness.

Notably, throughout the entire 200-epoch training cycle, there are no signs of overtraining or performance degradation on the validation set. This demonstrates that the proposed joint modeling mechanism, combining Diffusion and Transformer, achieves good optimization convergence in high-dimensional sparse data settings. Its ability to progressively refine semantic representation and structural reconstruction not only improves feature learning efficiency but also enhances model adaptability under complex data conditions.

4. Conclusion

This paper addresses the problem of deep mining in high-dimensional sparse feature spaces and proposes a data mining method based on the Diffusion-Transformer architecture. By introducing the stepwise perturbation and reconstruction mechanisms of diffusion models, the method effectively mitigates issues of information loss and weak feature expression caused by sparsity. Meanwhile, the Transformer structure plays a critical role in capturing global dependencies and extracting multi-scale semantics, enabling deep construction of latent structural relationships. high-dimensional textual features within diffusion models, significantly enhancing semantic restoration and relational inference. It serves as a major enhancement to traditional diffusion frameworks. This enhances the model's discriminative power and semantic consistency. Transformer provides global modeling capability for sparse

In the experimental section, the model is systematically evaluated on the public high-dimensional text dataset Reuters-21578 and compared against several mainstream deep mining methods. The results show that the proposed Diffusion-Transformer achieves leading performance in key metrics such as accuracy, precision, and recall. It also maintains strong stability under different sparsity levels. Additionally, the training and validation loss curves confirm the model's excellent optimization convergence and generalization ability. This study verifies the effective integration of diffusion modeling and attention structures in sparse high-dimensional data mining. However, the current model may still face limitations in computational complexity and efficiency, particularly in extremely high-dimensional and large-scale datasets. It also provides useful guidance for future model designs in tasks such as unstructured data processing, weak supervision, and anomaly pattern recognition. When handling problems like incomplete information and weak data correlation, the method demonstrates good adaptability to structure and strong capability in semantic recovery, laying a theoretical and practical foundation for building more generalizable mining models. Future research may further explore the adaptability of this framework in multimodal data environments, particularly in tasks like image-text fusion and alignment between sensor data and semantic labels [10]. Techniques such as self-supervised learning, knowledge graph guidance, and graph-based structure modeling can be integrated to enhance model performance under extremely sparse and noisy conditions [11]. These directions may help promote the deployment of intelligent mining methods in more complex real-world scenarios.

References

[1] Zaman, Ezzatul Akmal Kamaru, Azlinah Mohamed, and Azlin Ahmad. "Feature selection for online streaming high-dimensional data: A state-of-the-art review." Applied Soft Computing 127 (2022): 109355.

[2] Liu, Wenyi, et al. "A Recommendation Model Utilizing Separation Embedding and Self-Attention for Feature Mining." 2024 3rd International Conference on Cloud Computing, Big Data Application and Software Engineering (CBASE). IEEE, 2024.

[3] Zahariah, Siti, and Habshah Midi. "Minimum regularized covariance determinant and principal component analysis-based method for the identification of high leverage points in high dimensional sparse data." Journal of Applied Statistics 50.13 (2023): 2817-2835.

[4] Wang, Yongxin, et al. "A high-dimensional sparse hashing framework for cross-modal retrieval." IEEE Transactions on Circuits and Systems for Video Technology 32.12 (2022): 8822-8836.

[5] Robson, Barry, Srinidhi Boray, and J. Weisman. "Mining real-world high dimensional structured data in medicine and its use in decision support. Some different perspectives on unknowns, interdependency, and distinguishability." Computers in Biology and Medicine 141 (2022): 105118.

[6] Yoon, Jinsung, et al. "Vime: Extending the success of self-and semi-supervised learning to tabular domain." Advances in neural information processing systems 33 (2020): 11033-11043.

[7] Arik, Sercan Ö., and Tomas Pfister. "Tabnet: Attentive interpretable tabular learning." Proceedings of the AAAI conference on artificial intelligence. Vol. 35. No. 8. 2021.

[8] Wu, Haixu, et al. "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting." Advances in neural information processing systems 34 (2021): 22419-22430.

[9] Somepalli, G., Goldblum, M., Schwarzschild, A., Bruss, C. B., & Goldstein, T. (2021). Saint: Improved neural networks for tabular data via row attention and contrastive pre-training. arXiv preprint arXiv:2106.01342.

[10] Cheng, Q., Zhou, Y., Fu, P., Xu, Y., & Zhang, L. (2021). A deep semantic alignment network for the cross-modal image-text retrieval in remote sensing. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 4284-4297.

[11] Zhao, G., Zhang, C., Shang, H., Wang, Y., Zhu, L., & Qian, X. (2023). Generative label fused network for image–text matching. Knowledge-Based Systems, 263, 110280.