Mini Review

Feature Processing Methods: Recent Advances and Future Trends

Lufeng Bai[1]* and Shiying Bai[1,2]

[1]Jiangsu Second Normal College Nanjing, Jiangsu 211200, China
[2]Jiang Yan Second School, TaiZhou, 225500, China

Check for updates

Abstract

This paper reviews recent advances and future trends in feature processing methods within the field of artificial intelligence. With the rapid development of deep learning and big data technologies, feature processing has become essential for enhancing AI model performance. We begin by revisiting traditional feature processing methods, then focus on deep learning-based feature extraction techniques, automated feature engineering, and the application of feature processing in specific domains. The article also analyzes the current research challenges and outlines future development directions, offering structured insights for both researchers and practitioners across disciplines

Introduction

Feature processing constitutes a critical component in artificial intelligence and machine learning, directly impacting model performance and generalization capabilities. As data scales and complexity continue to increase, traditional methods often fall short in addressing the complexity of modern AI systems. Recent advances in deep learning have introduced novel opportunities for feature processing, positioning automated feature engineering as a key research focus. This paper aims to explore the latest advancements in AI feature processing methods, analyze current challenges, and prospect future trends to provide insights for related research.

Traditional feature processing methods

Data processing

Traditional feature processing methods mainly include data preprocessing, feature selection, and dimensionality reduction techniques. Data preprocessing is the first step in feature processing, including data cleaning, missing value processing, outlier detection, and data standardization. These steps are crucial for improving data quality and subsequent feature processing effectiveness. Data cleaning involves processing noisy data and identifying erroneous values, while missing value processing includes interpolation, deletion, or using algorithms to predict missing values. Outlier detection identifies outliers in data through statistical methods or machine learning algorithms to prevent their negative impact on model training. Data standardization uses normalization or normalization methods to transform features of different scales into the same range, in order to improve the convergence speed and performance of the model.

Feature selection

Feature selection is the process of selecting the most valuable subset of features from the original feature set. Common methods include filtering, packaging, and embedding. The filtering method evaluates the importance of features through statistical indicators such as chi square test, mutual information, etc. The packaging method combines the feature selection process with model training, iteratively selecting the optimal feature subset. The embedding rule automatically performs feature selection during model training, such as Lasso regression and decision tree algorithms. These methods each have their own advantages and disadvantages [1,2] and in practical applications, they need to be selected and combined based on specific problems and data characteristics.

Dimensionality reduction

Dimensionality reduction techniques reduce the number of features while retaining important information. Principal

Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are two widely used linear dimensionality reduction methods. PCA transforms the original features into a set of linearly independent principal components through orthogonal transformation, thereby achieving dimensionality reduction. LDA is a supervised dimensionality reduction method that seeks the optimal projection direction by maximizing inter class distance and minimizing intra class distance. In addition, nonlinear dimensionality reduction methods such as t-SNE and UMAP perform well in handling complex data structures and are widely used in visualization and high-dimensional data analysis.

These traditional methods have played an important role in past AI applications, but as data complexity increases, their limitations in handling non-linear relationships and automatic feature extraction become increasingly apparent. This has prompted researchers to explore more advanced feature processing methods based on deep learning.

Feature extraction based on deep learning

Feature extraction with CNNs

The rise of deep learning technology has brought revolutionary changes to feature processing. Convolutional Neural Networks (CNNs) perform well in the field of image processing, capable of automatically learning hierarchical feature representations. CNNs extract local features using convolutional layers, reduce dimensionality with pooling layers, and capture global patterns via fully connected layers , thereby achieving efficient feature extraction of images. In computer vision tasks, pre trained CNN models such as ResNet and Inception have become standard tools for feature extraction [3-5].

Sequence modeling via RNNs

Recurrent Neural Networks (RNNs) and their variants, such as Long Short Term Memory Networks (LSTM) [6-8], exhibit powerful feature extraction capabilities when processing sequential data. RNN captures temporal dependencies in sequence data through a cyclic structure, while LSTM effectively solves the problem of gradient vanishing in long sequence training by introducing memory units and gating mechanisms. These models have achieved significant results in fields such as natural language processing and speech recognition.

Autoencoders & generative models

The autoencoder, an unsupervised learning model, captures effective data representations through encoding and decoding phases , and has wide applications in feature dimensionality reduction and denoising. The basic autoencoder consists of an encoder and a decoder, which learn low dimensional representations of data by minimizing reconstruction errors. Extended models such as Variational

Autoencoder (VAE) and Generative Adversarial Network (GAN) [9-12] have further improved the expressive power of feature learning and performed well in tasks such as image generation and data augmentation.

These deep learning methods can automatically learn complex feature representations from raw data, greatly reducing the workload of manually designing features. However, they also face challenges such as poor interpretability and the need for large amounts of training data, which has driven the development of automated feature engineering.

Automated feature engineering and latest research hotspots

Neural architecture search

Automated feature engineering aims to reduce manual intervention and improve the efficiency and effectiveness of feature processing. Neural Architecture Search (NAS) optimizes the feature extraction process by automating the design of neural network structures [13-17]. NAS applies reinforcement learning, evolutionary algorithms, or gradient-based techniques to identify optimal architectures within a defined search space within a predefined search space. This approach enhances model performance while significantly reducing the burden of manual architecture design. For more details of NAS, (Figure 1).
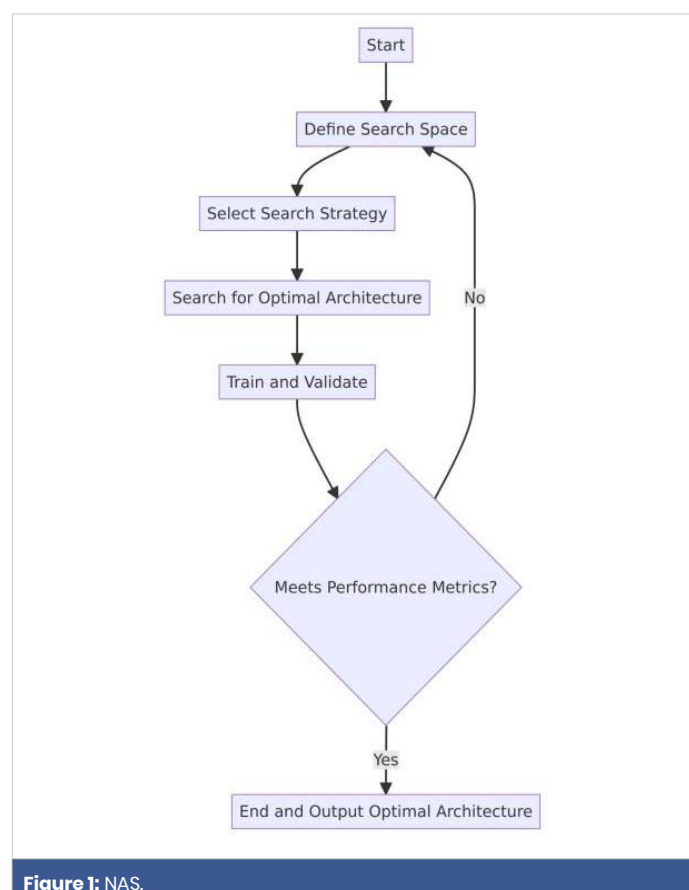

Figure 1: NAS.

Meta-learning for feature adaptation

Meta learning methods enable models to quickly adapt to the feature processing requirements of new tasks by learning how to learn. Meta learning trains models on multiple related tasks to enable them to quickly adapt to new tasks.

This method performs well in small sample learning and cross domain transfer learning, providing new ideas for feature processing.

RL-based feature selection framework

The feature selection method based on reinforcement learning dynamically optimizes the feature selection strategy through interaction with the environment. This method models the feature selection process as a Markov decision process, guiding the agent to learn the optimal feature subset through a reward mechanism. Reinforcement learning methods have unique advantages in dealing with high-dimensional data and dynamic feature selection problems [18-21].

These methods represent the forefront of current feature processing research [22,23], not only improving the automation level of feature processing, but also providing new ideas for processing high-dimensional and heterogeneous data [24]. However, how to balance the degree of automation with model interpretability, and how to effectively apply these methods to practical scenarios, is still a problem that needs further research [25,26].

The application and challenges of feature processing in specific fields

Feature processing methods have shown great potential in different AI application fields. In the field of computer vision, feature processing techniques have improved the accuracy of image recognition and object detection. For example, in medical image analysis, feature processing methods that combine domain knowledge can effectively extract lesion features and assist doctors in diagnosis. In the field of autonomous driving, multi-sensor data fusion and real-time feature processing technology are key to achieving environmental perception and decision-making.

Enhance the depth of algorithm application and cross modal practice

In natural language processing, word embeddings and context aware feature representations significantly improve the performance of language models. Pre trained language models such as BERT, GPT, etc. learn universal language features through large-scale corpora and perform well in downstream tasks. However, how to effectively handle multilingual and multimodal text data, as well as how to achieve effective feature representation in low resource languages, remains an urgent problem to be solved.

In the technical domain, integrating advanced algorithms with real-world applications can significantly enhance the practical relevance of feature processing. For instance, in industrial defect detection, convolutional neural networks (CNNs) coupled with attention mechanisms (e.g., CBAM [27]) have demonstrated superior performance in identifying micro-cracks on metal surfaces. A comparative study on the NEU-DET dataset revealed a 15% accuracy improvement over traditional methods like Haralick texture analysis. In autonomous driving, multimodal feature fusion techniques address the challenges of heterogeneous sensor data. The PointPainting algorithm [28] effectively aligns LiDAR point clouds with camera images by projecting semantic labels from 2D images onto 3D points, achieving a 12% higher mAP on the nuScenes benchmark. Additionally, cross-domain adaptation of architectures, such as applying Transformer-based models (e.g., ESM-2 [29]) to protein sequence analysis, illustrates how NLP-inspired feature representations can advance bioinformatics tasks like fold recognition. These examples highlight the need to tailor deep learning frameworks to domain-specific challenges and promote interdisciplinary integration while leveraging interdisciplinary insights.

Deepen medical data-driven solutions

In the field of bioinformatics, effective feature processing is crucial for gene sequence analysis and protein structure prediction. Deep learning models such as AlphaFold have achieved breakthroughs in protein structure prediction by combining evolutionary information and physical constraints [30]. However, the complexity, high dimensionality, and noise issues of biological data pose significant challenges to feature processing, requiring the development of more robust and interpretable feature processing methods.

Clinical applications require robust feature processing methods that optimize accuracy, ensure interpretability, and preserve data privacy. In breast cancer histopathology, Vision Transformers (ViTs) have outperformed traditional texture-based features (e.g., Haralick descriptors) by achieving an AUC of 0.92 versus 0.85 on the BreakHis dataset [31], attributed to their ability to capture global contextual patterns. To mitigate data scarcity, federated learning frameworks support multi-institutional collaboration without compromising patient privacy. For example, a federated feature extraction model trained on mammography data from five hospitals improved malignancy detection F1-scores by 18% compared to singlecenter models [32]. Furthermore, interpretability tools like Grad-CAM [33] bridge the gap between AI decisions and clinical trust. In lung nodule detection, Grad-CAM heatmaps highlight malignancy-associated regions in CT scans , aligning with radiologists diagnostic criteria (validated in a 2023 Nature Medicine study [34]). These advances emphasize the need for feature engineering that harmonizes technical innovation with clinical workflows and ethical considerations.

Addressing the challenges of robustness and real-time performance in industrial scenarios

Engineering applications require feature processing methods that address realworld robustness and efficiency constraints. In semiconductor wafer inspection, self-supervised learning frameworks like SimCLR [35] extract discriminative features from limited labeled data, reducing false defect rates from 8% to 3% on the WM-811K dataset. For real-time systems, lightweight architectures such as MobileNetV3 [36] optimize feature extraction on edge devices. Deployed on NVIDIA Jetson Xavier, MobileNetV3 reduced inference latency from 50 ms to 20 ms while maintaining 98% accuracy in automotive object detection. Multimodal fusion also plays a pivotal role in predictive maintenance: integrating vibration sensors, thermal imaging, and acoustic signals via hybrid models (e.g., wavelet-CNN [37] achieved 96% fault prediction accuracy in a smart factory case study (IEEE Transactions on Industrial Informatics, 2023 [38]). These use cases emphasize balancing computational efficiency, data heterogeneity, and real-world variability in industrial AI applications.

These applications also face unique challenges. For example, how to handle feature fusion of multimodal data, how to deal with data scarcity and class imbalance, and how to improve the interpretability of the feature processing process. Addressing these challenges requires interdisciplinary collaboration and innovative thinking.

Conclusion

As a core component of artificial intelligence, feature processing methods are undergoing rapid development and transformation. From traditional methods to deep learning based automatic feature extraction, and then to automated feature engineering, new technologies and methods are constantly emerging in this field. In the future, feature processing research may pay more attention to multimodal data fusion, small sample learning, interpretability, and other aspects. As AI application scenarios expand, the development of more universal and efficient feature processing methods is expected to become a key focus. These advancements are expected to elevate AI development and support innovative applications across domains.

References

1. Dhal P, Azad C. A comprehensive survey on feature selection in the various fields of machine learning. Appl Intell. 2022;52(4):4543-4581. Available from: https://link.springer.com/article/10.1007/s10489-021-02550-9

2. Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. Nat Med. 2022;28(9):1773-1784. Available from: https://doi.org/10.1038/s41591-022-01981-2

3. Alotaibi B, Alotaibi M. A hybrid deep ResNet and inception model for hyperspectral image classification. PFGC J Photogramm Remote Sens Geoinf Sci. 2020;88(6):463-476. Available from: https://link.springer.com/article/10.1007/s41064-020-00124-x

4. Peng S, Huang H, Chen W, Zhang L, Fang W. More trainable inception-ResNet for face recognition. Neurocomputing. 2020;411:9-19. Available from: https://doi.org/10.1016/j.neucom.2020.05.022

5. Barakbayeva T, Demirci FM. Fully automatic CNN design with inception and ResNet blocks. Neural Comput Appl. 2023;35(2):1569-1580. Available from: http://dx.doi.org/10.1007/s00521-022-07700-9

6. Sherstinsky A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D Nonlinear Phenom. 2020;404:132306. Available from: https://doi.org/10.1016/j.physd.2019.132306

7. Al-Selwi SM, Hassan MF, Abdulkadir SJ, Muneer A, Sumiea EH, Alqushaibi A, et al. RNN-LSTM: From applications to modeling techniques and beyond—Systematic review. J King Saud Univ Comput Inf Sci. 2024:102068. Available from: https://doi.org/10.1016/j.jksuci.2024.102068

8. Shewalkar A, Nyavanandi D, Ludwig SA. Performance evaluation of deep neural networks applied to speech recognition: RNN, LSTM and GRU. J Artif Intell Soft Comput Res. 2019;9:235-245. Available from: http://dx.doi.org/10.2478/jaiscr-2019-0006

9. Gao R, Hou X, Qin J, Chen J, Liu L, Zhu F, et al. Zero-VAE-GAN: Generating unseen features for generalized and transductive zero-shot learning. IEEE Trans Image Process. 2020;29:3665-3680. Available from: https://doi.org/10.1109/tip.2020.2964429

10. Tian C, Ma Y, Cammon J, Fang F, Zhang Y, Meng M. Dual-encoder VAE-GAN with spatiotemporal features for emotional EEG data augmentation. IEEE Trans Neural Syst Rehabil Eng. 2023;31:2018-2027. Available from: https://doi.org/10.1109/tnsre.2023.3266810

11. Ibrahim BI, Nicolae DC, Khan A, Ali SI, Khattak A. VAE-GAN based zero-shot outlier detection. In: Proceedings of the 2020 4th international symposium on computer science and intelligent control. 2020. Available from: https://doi.org/10.1145/3440084.3441180

12. Mukesh K, Ippatapu VS, Chereddy S, Anbazhagan E, Oviya IR. A variational autoencoder general adversarial networks (VAE-GAN) based model for ligand designing. In: International Conference on Innovative Computing and Communications: Proceedings of ICICC 2022, Volume 1. Singapore: Springer Nature Singapore; 2022. Available from: https://www.amrita.edu/publication/a-variationalautoencoder-general-adversarial-networks-vae-gan-based-model-for-ligand-designing/

13. Elsken T, Metzen JH, Hutter F. Neural architecture search: A survey. J Mach Learn Res. 2019;20(55):1-21. Available from: https://www.jmlr.org/papers/volume20/18-598/18-598.pdf

14. Ren P, Xiao Y, Chang X, Huang PY, Li Z, Chen X, et al. A comprehensive survey of neural architecture search: Challenges and solutions. ACM Comput Surv. 2021;54(4):1-34. Available from: https://arxiv.org/abs/2006.02903

15. Chitty-Venkata KT, Somani AK. Neural architecture search survey: A hardware perspective. ACM Comput Surv. 2022;55(4):1-36. Available from: http://dx.doi.org/10.1145/3524500

16. Li L, Talwalkar A. Random search and reproducibility for neural architecture search. In: Uncertainty in artificial intelligence. PMLR; 2020. p. 367-377. Available from: https://arxiv.org/abs/1902.07638

17. Lindauer M, Hutter F. Best practices for scientific research on neural architecture search. J Mach Learn Res. 2020;21(243):1-18. Available from: https://doi.org/10.48550/arXiv.1909.02453

18. Mousavi SS, Schukat M, Howley E. Deep reinforcement learning: an overview. In: Proceedings of SAI Intelligent Systems Conference (IntelliSys) 2016: Volume 2. Springer International Publishing; 2018. Available from: https://doi.org/10.48550/arXiv.1806.08894

19. Ding Z, Huang Y, Yuan H, Dong H. Introduction to reinforcement learning. Deep reinforcement learning: fundamentals, research and applications. 2020:47-123. Available from: http://dx.doi.org/10.1007/978-981-15-4095-0_2

20. Mosavi A, Faghan Y, Ghamisi P, Duan P, Ardabili SF, Salwana E, et al. Comprehensive review of deep reinforcement learning methods and applications in economics. Mathematics. 2020;8(10):1640. Available from: https://doi.org/10.3390/math8101640

21. Barto AG. Reinforcement learning: An introduction. SIAM Rev. 2021;6(2):423.

22. Gahar RM, Arfaoui O, Hidri MS, Hadj-Alouane NB. A distributed approach for high-dimensionality heterogeneous data reduction. IEEE Access. 2019;7:151006–151022. Available from: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8861035

23. Yilmaz Y, Aktukmak M, Hero AO. Multimodal data fusion in high-dimensional heterogeneous datasets via generative models. IEEE Trans Signal Process. 2021;69:5175–5188. Available from: https://doi.org/10.48550/arXiv.2108.12445

24. Pölsterl S, Conjeti S, Navab N, Katouzian A. Survival analysis for high-dimensional, heterogeneous medical data: Exploring feature extraction as an alternative to feature selection. Artif Intell Med. 2016;72:1–11. Available from: https://doi.org/10.1016/j.artmed.2016.07.004

25. Rabiee M, Mirhashemi M, Pangburn MS, Piri S, Delen D. Towards explainable artificial intelligence through expert-augmented supervised feature selection. Decis Support Syst. 2024;181:114214. Available from: https://doi.org/10.1016/j.dss.2024.114214

26. Aguilar-Ruiz JS. Class-specific feature selection for classification explainability. ArXiv Preprint. 2024. Available from: https://doi.org/10.48550/arXiv.2411.01204

27. Woo S, Park J, Lee JY, Kweon IS. CBAM: Convolutional block attention module. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2018. Available from: https://doi.org/10.48550/arXiv.1807.06521

28. Vora S, Lang AH, Helou B, Beijbom O. PointPainting: Sequential fusion for 3D object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020. p. 4604–4612. Available from: https://doi.org/10.48550/arXiv.1911.10150

29. Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. Science. 2023;379:1123-1130. Available from: https://doi.org/10.1126/science.ade2574

30. Rostami M, Oussalah M. A novel explainable COVID-19 diagnosis method by integration of feature selection with random forest. Inform Med Unlocked. 2022;30:100941. Available from: https://doi.org/10.1016/j.imu.2022.100941

31. Panhol FA, Oliveira LS, Petitjean C, Heutte L. A dataset for breast cancer histopathological image classification. IEEE Trans Biomed Eng. 2016;63(7):1455-1462. Available from: https://doi.org/10.1109/tbme.2015.2496264

32. Li X. Federated feature learning for mammography diagnosis with privacy preservation. Nat Digit Med. 2022;5:123.

33. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. Proc IEEE Int Conf Comput Vis (ICCV). 2017;618-626. Available from: https://doi.org/10.48550/arXiv.1610.02391

34. Wang H. Interpretable AI for lung nodule malignancy prediction in low-dose CT. Nat Med. 2023;29(6):1430-1438.

35. Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. Proc Int Conf Mach Learn (ICML). 2020;119:1597-1607. Available from: https://proceedings.mlr.press/v119/chen20j.html

36. Howard A, Sandler M, Chu G, Chen LC, Chen B, Tan M, et al. Searching for MobileNetV3. Proc IEEE/CVF Int Conf Comput Vis (ICCV). 2019;1314-1324. Available from: https://openaccess.thecvf.com/content_ICCV_2019/html/Howard_Searching_for_MobileNetV3_ICCV_2019_paper.html

37. Zhang Y. Wavelet-CNN for mechanical fault diagnosis under noisy environments. Mech Syst Signal Process. 2021;152:107413.

38. Gupta A. Multimodal sensor fusion for predictive maintenance in Industry 4.0. IEEE Trans Ind Inform. 2023;19(7):4321-4332.

39. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583-589. Available from: https://www.nature.com/articles/s41586-021-03819-2