Trajectory pattern extraction to climate data and climate change analytics

Xin Zhang[1], Jingyue Zhang[2], Pingping Wang[3]

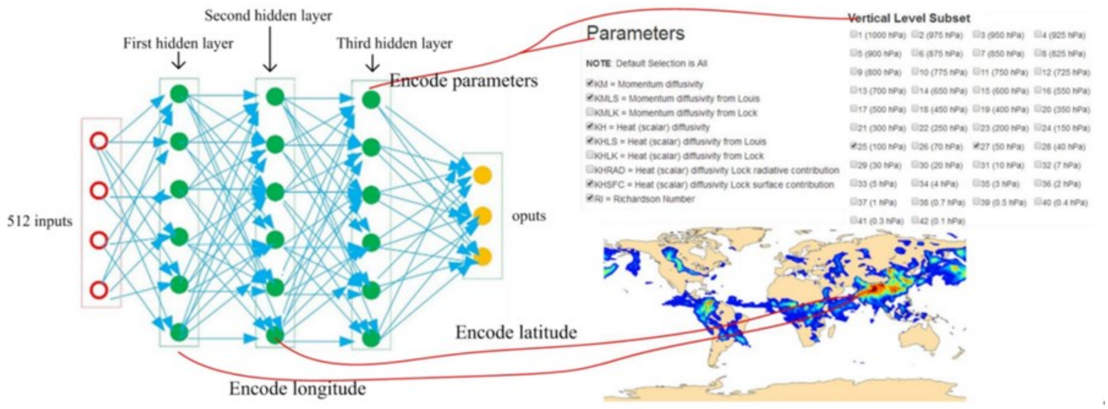
[1,3] School of Intelligent Science and Engineering, Yunnan Technology and Business University, 651701, Kunming, China.

[2] College of Big Data, China Polytechnic Institute, 401320, China

Corresponding author: Xin Zhang.

E-mail: zhangxingyunnan@163.com

Graphical abstract



Abstract:

Climate data contains rich information and show high dimensions. Through exploring the data trajectory of climate data, climate change can be forecasted. Climate data is typical high-dimensional data, unlike shallow structures, deep structures have more advantages in handling high-dimensional data. Deep neural networks with multiple layers have ability of the data trajectory discovery from the complicated high-dimensional data. Here, this paper seeks to explore the data trajectory from complicated climate data using the proposed deep neural networks. First, using the sampling theorem to reconstruct the discrete surface that can converge to the original space. Thereafter, to minimize the error between the original data trajectory and the extracted data trajectory, the Brenier theoretical function in Brenier Theorem is used as the loss function of the deep neural networks. By doing so, the deep neural networks can extract the data trajectory from climate data as accurate as possible. Finally, as an important application, since a thorough understanding of the data trajectory of climate data is essential for understanding the effect of earth's future climate.

Keyword: Climate data, trajectory

1. Introduction

Climate data has obvious spatial and temporal characteristics, which not only requires real-time updates but also needs precise measurements at different times and locations. Additionally, climate data is also influenced by these factors, such as terrain and altitude, thereby displaying complex characteristics for the data on spatial distribution. Therefore, climate data has both regular temporal characteristics and obvious spatiality. Climate data contains rich valuable information, through discovering the data trajectory, climate change can be forecasted. Due to climate data is a type of high-dimensional data, intelligent methods based on automatic extraction relation patterns are popular [1]. For example, deep learning has some applications in terms of automatic extraction relation patterns like seeing, listening and speaking [2][3]. Deep neural networks can capture as accurately as possible the correct input during training sample [4], which may be constructed from many-layered, i.e., ‘deep’. Deep neural networks have been used to solve difficult problems in the field of image and speech recognition [5]. The Ref. [6] and Ref. [7] indicate that the stacked neural networks can extract complicated patterns from high-dimensional data.

Recently, deep neural networks have been used for extracting the trajectory from the complicated high-dimensional data. Goldberg et al [8] indicate that the trajectory learning is based on that the data samples or their features reside a low-dimensional trajectory. Bethany et al. [9] succeeded in combining the linear embedding and the auto-encoder to recognize the nonlinear coordinates, however, the approach needs to dynamical adjust. Tompson et al. [10] propose that it is effectively to solve the complex machine learning problems by introducing the ‘weak learners’ into an ensemble approach to develop relatively simple solutions, but the method gets some simple solutions.

Motivation**.** In this brief, we study how to extract the data trajectory from the complicated climate data containing complicated multiple variables. To finish the goal, we proposed the multiple-layer deep neural networks. On the hand, we extract the data point from the original of climate data with the dense sampling, because sampling theorem enables that the reconstruction discrete surface can converge to the original surface of climate data. On the other hand, to minimize the error between the original data trajectory and the extracted data trajectory, the loss function was derived using Brenier’s theorem. By doing so, the deep neural networks can extract the data trajectory from climate data as accurately as possible.

The rest of the paper is organized as follows. Section 2 summarized the related works. In Section 3, we used the sampling theorem to extract the data pairs, then introduce the Brenier theoretical function into the deep neural networks as the loss function. Section 4 verified the proposed method, and discussed the experimental results. Finally, Section 5 drew the conclusion and directed the future work.

2. Related works

Scholars proposed the thoughts for data trajectory discovery and patterns extraction, including (i) Feature-based methods, which represent high-dimensional data through choosing a small feature subset [11][12]. Such as the method based on domain knowledge feature proposed in Ref. [13] and Ref. [14]. And the method based on feature coupling generalization [15]. Feature-based methods can save rich information as far as possible and filter irrelevant attributions [16][17], moreover, the obtained information is interpretable. Currently, Feature-based methods can be categorized into filter method, wrapper method, embedded method. Filter method evaluates feature scores through using data intrinsic features [18], however, there might obtain poor patterns. Such as the method implemented in Ref. [19] and Ref. [20]. The method [21] based on wrapper shows superior performance, while encounters a risk of falling into local optimum. Compared to wrapper method, embedded method, such as the method proposed in Ref. [22], has relatively lower cost consumption [23].

(ii) Kernel-based methods, which can capture valuable information from background space [24]. The selection of the kernels has a critical effect on the results for such methods. To explore deeper relation patterns, Wei et al. [25] took account into a multiple kernel rather than a single kernel. Similarly, these instances implemented in Ref. [26-28].

(iii) Deep architectures-based methods, which have natural ascendency to capture deep representations between the data [29]. Such as Recurrent Neural Networks (RNNs) [30] and the Deep Convolutional Neural Networks [31]. The Ref. [32] uses the deep neural networks to successfully capture the trajectorys from the high-dimensional data, and indicates that high dimensionality of the data has a negative impact on the trajectorys. Moreover, extremely high dimensions of the data may lead to the extracted trajectory deformity.

3. Methodology

In terms of deep learning, each point of the trajectory can be mapped to the point of the parameter domain, that is, points in the background space are mapped to points in the hidden space, called encoding. The deep neural networks map a neighborhood of data trajectory to the hidden space and the probability distribution of the data to the parameter domain in the hidden space. Different coding mappings present different probability distributions in the hidden space in the process of analyzing data.

3.1 Sampling data

We apply dense sampling on the data space, then infinitely approximate the original space by the sampling data. The approximation accuracy is determined by the sampling density. Based on this, the (δ, ε) net is considered, and the sampling condition should satisfy that there is at least one sampling point inside any geodesic disk of radius δ on the surface, and that the distance between any two sampling points is no less than ε . This is to ensure that the reconstructed discrete surface can converge to the original surface. The common convergence distances contain Hausdorff distance, geodesic distance, curvature measure and Laplace-Beltrami operator and so on. Here, we chose the Hausdorff distance.

3.2 The model

Lemma 1. Brenier theorem [33]. Suppose x and y are the Euclidean space $\mathbb{R}^n$, and the transportation cost is the quadratic Euclidean distance $c(x, y) = |x - y|^2$. If u is absolutely continuous and μ , ν have finite second order moments, then there exists a convex function $\mu : X \to \mathbb{R}$, such that the gradient map $\nabla\mu$ gives the unique solution to the Monge's problem, where μ is called Brenier's potential, $\nabla\mu$ is called Brenier mapping or the optimal transmission mass mapping. In general, μ is not unique.

According to the Brenier theorem, if the source probability measure u satisfies some very broad conditions, such as absolute continuity, the finite of second moment, then the optimal transmission map exists and is unique. If the optimal transmission map is a gradient map of a convex function, then the convex function is called the Brenier potential energy, i.e., $\mu : \Omega \to \mathbb{R}$, where Ω is its convex area. The function μ is convex, that is, the Hessian matrix of the function is positive definite.

The gradient mapping of Brenier's potential function is $\nabla\mu : \Omega \to \Omega$. Using the gradient mapping can map point $p \in \Omega$ to $\nabla\mu(p) \in \Omega$. So, the gradient mapping should satisfy the Jacobi equation, and from that, we get the Monge-Ampere equation, having that

$$\det(D^2\mu) = \frac{\mu}{\nu \circ \nabla\mu} \tag{1}$$

The optimal transfer mapping [34] is the gradient of the convex function u, i.e., $f = \nabla\mu$. Gu et al. [35] proved the existence of solution of Monge-Ampere equation, then proposed that it solved the solution with convex geometry. We added the Brenier function into the deep neural networks, that is, we modified the loss function of deep neural networks as follows.

$$\begin{cases} J_*(W,b) = J(W,b) + \Theta \\ J(W,b) = \dfrac{1}{n}\sum_{i=1}^{n} J(W,b:x^{(i)},y^{(i)}) + \sum_{k=1}\sum_{i=1}\sum_{j=1}(W_{ji}^{k})^2 \end{cases} \tag{2}$$

Where Θ is the Brenier function. The gradient descent method is used to solve the optimization parameters, as shown in algorithm 1.

Fig. 1 unveils that the deep neural networks encoded the climate data. The deep neural networks contain three hidden layers, then each hidden layer encodes the different contents. Our objects are that we find low-dimensional (3-dimension) representation of each sample to discovering trajectory from the 8 variables. The first layer and second layer encoded the longitude and latitude of location, respectively, then the third layer encoded the climate parameters. Date from, 2015 (360 days). Low spatial (earth surface 80km-120km), 8 latitudes, 8 longitudes. Each sample composed of 8 × 8 × 8 = 512 variables.

Time complexity. The time consumption of Algorithm 1 consists of the training time of the model. Assuming that the model is trained n time, and the data dimensionality is m, therefore, the time complex is $O(n*m)$

Algorithm 1.

Initialization $W^{(k)} = 0$ and $b^{(k)} = 0$

For i=1 to n:

Using back propagation to calculate $\nabla_{W^{(k)}} J(W,b:x,y), \nabla_{b^{(k)}} J(W,b:x,y)$;

Setting $W^{(k)} = W^{(k)} + \nabla_{W^{(k)}} J(W,b:x,y)$;

Setting $b^{(k)} = b^{(k)} + \nabla_{b^{(k)}} J(W,b:x,y)$;

Updating the parameters:

$$W^{(k)} = W^{(k)} - \alpha[\frac{1}{m}W^{(k)} + \lambda W^{(k)}];$$

$$b^{(k)} = b^{(k)} - \alpha[\frac{1}{m}b^{(k)}];$$

End For

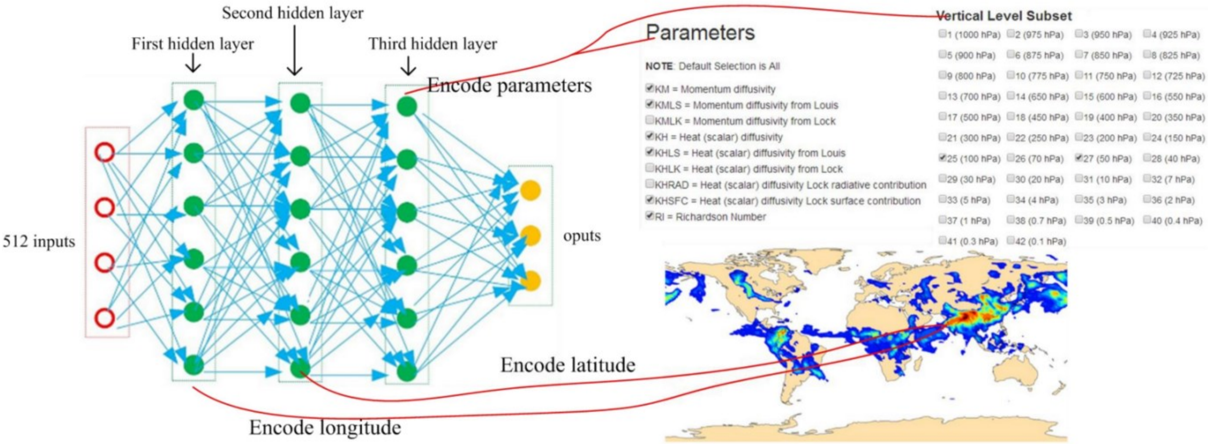


Fig. 1 Encoding the climate data. The first layer and second layer encoded the longitude and latitude of location, respectively, then the third layer encoded the climate parameters.

4. Experiment results

We applied the ECMWF climate data to verify the proposed method. ECMWF climate data in a year is used as experimental dataset. Data dimensionality and data volume are 512 and 40000. For the description of the dataset, please refer at https://atmosphere. copernicus.eu/

Apart from our method, the Kernel Principal components analysis (KPCA) [28] and Deep CNN [31] are used as a comparison.

4.1 Performance analytic

Fig. 2 showed this results that our deep neural networks applied the different function as model loss function. We compared Brenier function (BF) with Square loss function (SLF), Absolute loss function (ALS) and Exponential loss function (ELS). Compared results show that the loss value in Fig. 2 (*a*) is lower than that in Fig. 2 (*b*), Fig. 2 (*c*) and Fig. 2 (*d*). The final loss value of BF is 0.03, and the loss value of three loss functions SLF, ALS, ELS is 0.1. In the process of testing, the three loss functions LF, ALS, ELS present high frequency fluctuation. Additionally, BF presented better stability than the three loss functions LF, ALS, ELS, that is, it converges faster than them. That is why we chose the Brenier function as our loss function of the deep neural networks.

Fig. 3 presented the reconstruction error with three different approaches, displaying that the smallest and biggest mean value of reconstruction error are our deep neural networks and KPCA, respectively. Because of introducing the Brenier potential energy function into our deep neural network, we obtain the lower reconstruction error than the two opponents KPCA, Deep CNN. This means that the reconstruction error has an impact on extracting the trajectory of high-dimensional data.

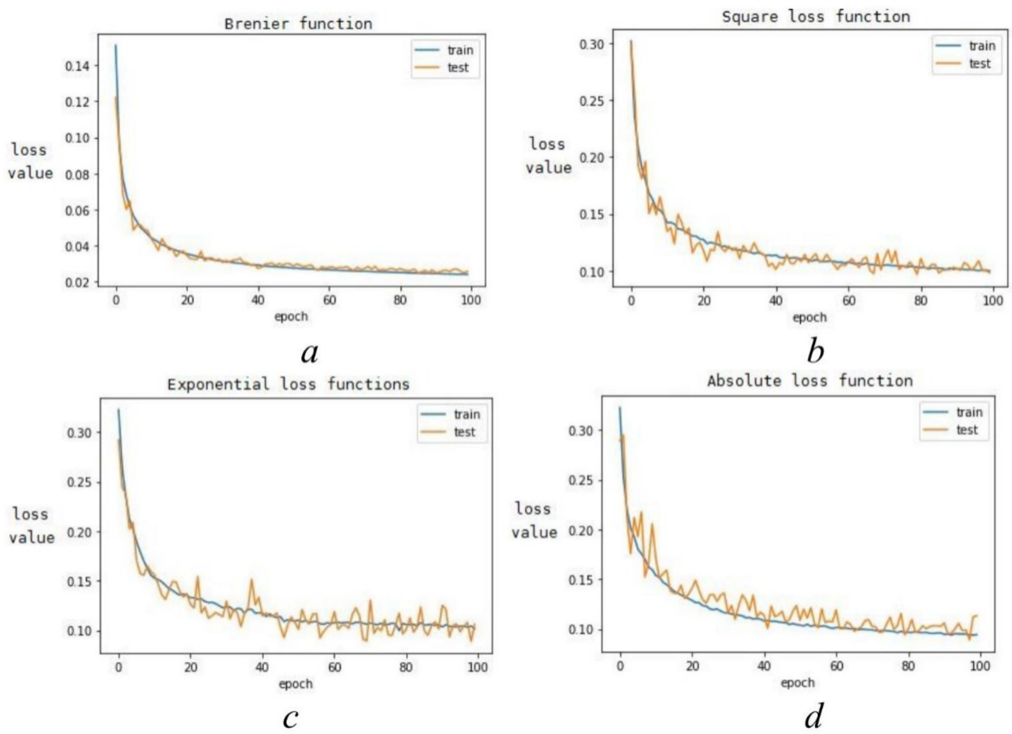


Fig. 2 Comparisons of different loss functions. (*a*), (*b*), (*c*), (*d*) display Brenier function, Square function, Absolute function and Exponential function as the loss function, respectively.

Our deep neural networks

```
reconstruction_error:
mean        0.035176
std         0.034309
min         0.000070
25%         0.000070
50%         0.034213
75%         0.064049
max         0.152274
```

Deep CNN

```
reconstruction_error:
Mean value: 0.119166
Iteration 10: error is 2.368928
Iteration 20: error is 2.328097
Iteration 30: error is 2.250574
Iteration 40: error is 2.239031
Iteration 50: error is 2.216933
Iteration 60: error is 2.190203
Iteration 70: error is 2.148016
Iteration 80: error is 2.087764
Iteration 90: error is 2.017473
Iteration 100: error is 1.949395
```

Kernel PCA

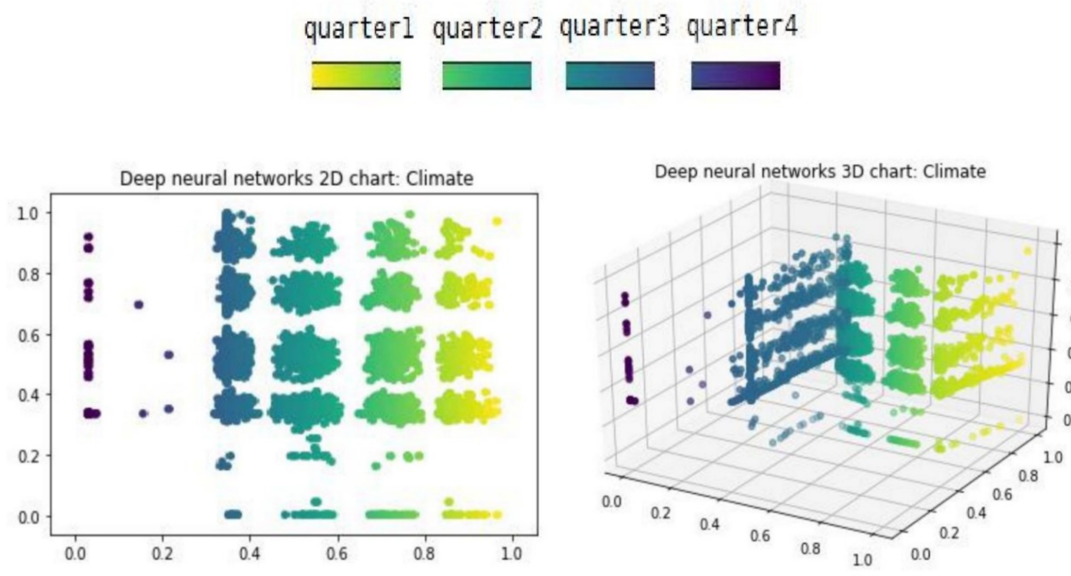
```
reconstruction_error:
Mean avlue: 0.245670
min    0.117352
max    0.373988
```

Fig. 3 Reconstruction error of the different approaches. The error comparisons between the three methods.

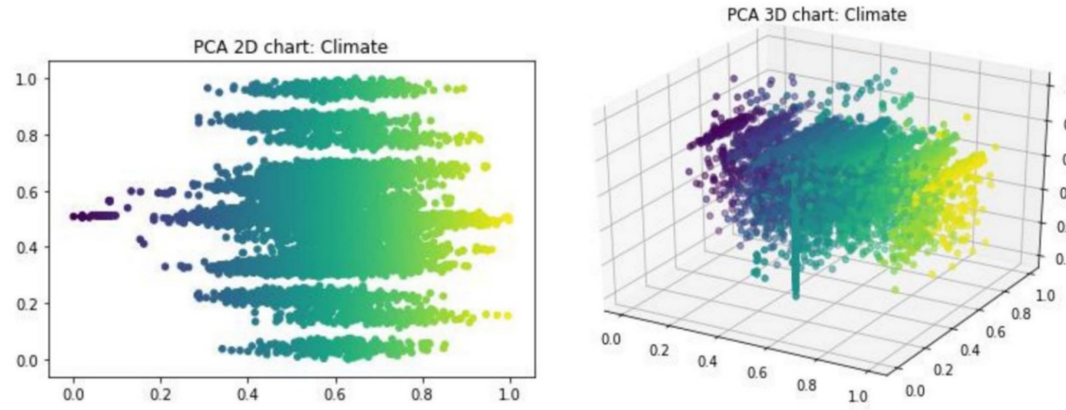
4.2 Comparisons of the extracted trajectory

The visualized results in Fig. 4 shows that the clustering results obtained by our deep neural networks are superior to those obtained by Deep CNN and KPCA. This means that our deep neural networks discover the data trajectory of climate data from the complicated climate data better than the two opponents Deep CNN and KPCA. In Fig. 4, it can be observed that the climate trajectory trajectories obtained by our model are more regular than those obtained by Deep CNN and by KPCA, because we applied the Brenier theorem and dense sampling in the process of analyzing the original data. Our deep neural networks contain three layers, as expected, which has captured the annual cycle. Through removing this non-eigenvalue information from the original data, we success in extracting the climate trajectory shape of each season.

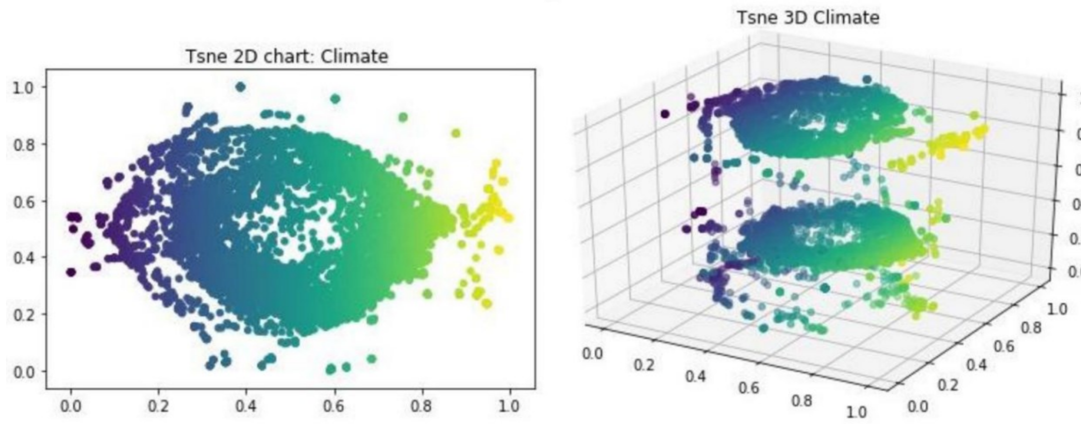
The reconstruction error has impact on the loss function, which means that we should consider the type of loss functions for extracting the trajectory from the complicated data with deep learning methods. Brenier theorem indicates that the Brenier function is the optimal transport mapping under certain conditions, so our deep neural networks gain the lowest loss value and better stability. Comparison with non-deep learning methods, the deep learning has more advantage in discovering the trajectory from the complicated high-dimensional data. The accuracy of the trajectory discovery is enhanced after introducing Brenier function into our deep neural network. Because in the process of adjusting weight, deep learning model makes the generated data distribution (output) to approximate the original data distribution (input).



a



b



c

Fig. 4. Visualizations of the ECMWF climate data. (a), (b), (c) display the deep neural networks visualization with 2-dimension and 3-dimension, KPCA visualization with 2-dimension and 3-dimension, Deep CNN visualization with 2-dimension and 3-dimension, respectively.

5. Conclusion

This paper investigated the problem the data trajectory discovery from the complicated climate data. We applied the sampling theorem to sample the original data, and then the deep neural networks with Brenier function are used as climate trajectory discovery. The experimental results show that the trajectory extracted by our deep neural networks are better than those extracted by the opponents. Furthermore, our reconstruction error is the smallest, compared with the opponents. In the future, we will apply deep forest to extract the complicated pattern from the climate data subject to noise environments. Noise can interfere the relation pattern extraction through mislead extractors.

Declaration

All authors have no conflicts of interest to declare that are relevant to the content of this article. All authors agree with availability of data and material in this work.

Data availability

Data will be made available on request. https://atmosphere.copernicus.eu/

References

[1] Jian Zheng, Qingling Wang, Cong Liu, et al. Relation patterns extraction from high-dimensional climate data with complicated multi-variables using deep neural networks. Applied Intelligence, 2023, 53:3124-3135.

[2] Chong Wang, Yuanhong Chen, Fengbei Liu, et al. An Interpretable and Accurate Deep-Learning Diagnosis Framework Modeled With Fully and Semi-Supervised Reciprocal Learning. IEEE Transactions on Medical Imaging, 2024, 43(1):392-404.

[3] Soyiba Jawed, Ibrahima Faye, Aamir Saeed Malik. Deep Learning-Based Assessment Model for Real-Time Identification of Visual Learners Using Raw EEG. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2024, 32:378-390.

[4] Jianzhu Ma, Michael Ku Yu , Samson Fong , et al. Using deep learning to model the hierarchical structure and function of a cell. Nature Method, 2018, 15:290-298.

[5] Vasileios Skaramagkas, Anastasia Pentari, Zinovia Kefalopoulou, et al. Multi-Modal Deep Learning Diagnosis of Parkinson's Disease—A Systematic Review. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2023, 31:2399-2423.

[6] Yang Yang, Yu Yan, Zhipeng Gao, et al. A Network Traffic Classification Method Based on Dual-Mode Feature Extraction and Hybrid Neural Networks. IEEE Transactions on Network and Service Management, 2023, 20(4): 4073-4084

[7] Weichao Lan, Yiu-Ming Cheung, Juyong Jiang, et al. Compact Neural Network via Stacking Hybrid Units. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(1):103-116.

[8] Y. Goldberg, A. Zakai, D. Kushnir, et al. Trajectory learning: The price of normalization. J. Mach. Learn. Res, 2018, 9:1909-1939.

[9] Bethany Lusch, J. Nathan Kutz, Steven L. Brunton, Deep learning for universal linear embeddings of nonlinear dynamics. Nature Communications, 2018, 9(1):1-14.

[10] J. Tompson, M. Stein, Y. Le Cun, et al. Real-time continuous pose recovery of human hands using convolutional networks. ACM Transactions on Graphics, 2014, 33(169):1-1.

[11] M. Yamada, J. Tang, J. Lugo-Martinez, et al. Ultra high-dimensional nonlinear feature selection for big biological data. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(7):1352-1365.

[12] T. Pang, F. Nie, J. Han, et al. Efficient feature selection via l 2,0-norm constrained sparse regression. IEEE Transactions on Knowledge and Data Engineering, 2018, 3(5):880-893.

[13] Lei Wang, Jinhai Liu, Huaguang Zhang, et al. KMSA-Net: A Knowledge-Mining-Based Semantic-Aware Network for Cross-Domain Industrial Process Fault Diagnosis. IEEE Transactions on Industrial Informatics, 2024, 20(2):2738 -2750.

[14] Binhui Xie, Shuang Li, Fangrui Lv, et al. A Collaborative Alignment Framework of Transferable Knowledge Extraction for Unsupervised Domain Adaptation. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(7):6518-6533.

[15] Run Jiang, Guanghai Bao, Qiteng Hong, et al. Machine Learning Approach to Detect Arc Faults Based on Regular Coupling Features. IEEE Transactions on Industrial Informatics, 2023, 19(3):2761-2771.

[16] J. Berón, H. D. B. Restrepo, A. C. Bovik. Optimal feature selection for blind super-resolution image quality evaluation. In IEEE International Conference on Acoustics, Speech and Signal Processing, 2019 IEEE, pp.1842-1846.

[17] S. Y. Sekeh, A. O. Hero. Feature selection for mutlti-labeled variables via dependency maximization. In IEEE International Conference on Acoustics, Speech and Signal Processing, 2019, IEEE, pp.3127-3131.

[18] R. Zhang, F. Nie, X. Li, et al. Feature selection with multi-view data: A survey. Information Fusion, 2019, 50:158-167.

[19] M. Mafarja, S. Mirjalili. Whale optimization approaches for wrapper feature selection. Applied Soft Computing, 2018, 62:441-453.

[20] O. Gokalp, E. Tasci, A. Ugur. A novel wrapper feature selection algorithm based on iterated greedy metaheuristic for sentiment classification. Expert Systems with Applications, 2020, 146: p.113176.

[21] Learning Wrapper Techniques, Mojtaba Yavandhasani, Foad Ghaderi. Visual Object Recognition From Single-Trial EEG Signals Using Machine. IEEE Transactions on Biomedical Engineering, 2022, 69(7):2176-2183.

[22] He Li, Yao Zhang, Yang Yang, et al. Performance Analysis of Coal Gangue Recognition Based on Hierarchical Filtering and Coupled Wrapper Feature Selection Method. IEEE Access, 2023, 11:85822-85835.

[23] 16. X. Li, H. Zhang, R. Zhang, et al. Generalized uncorrelated regression with adaptive graph for unsupervised feature selection. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(5):1587-1595.

[24] Zhao Zhang, Lei Jia, Mingbo Zhao. Kernel-Induced Label Propagation by Mapping for Semi-Supervised Classification. IEEE Transactions on Big Data, 2019, 5(2):148-165.

[25] Wei Huang, Yao Huang, Hua Wang. Local Binary Patterns and Superpixel-Based Multiple Kernels for Hyperspectral Image Classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing , 2020, 13:4550-4563.

[26] Jiyuan Liu, Xinwang Liu, Jian Xiong, et al. Optimal Neighborhood Multiple Kernel Clustering With Adaptive Local Kernels. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(6):2872-2885.

[27] Kenneth Yeonkong Ma, Chein-I. Chang. Kernel-Based Constrained Energy Minimization for Hyperspectral Mixed Pixel Classification. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60:1-12.

[28] Zhaoyu Lei, Jianyi Guo, Feng Zheng, et al. Thyristor State Evaluation Method Based on Kernel Principal Component Analysis. IEEE Access, 2022, 10:29992-30004.

[29] Jian Zheng, Xin Hu. irrelevant attribute resistance approach to binary classification for imbalanced data. Information Sciences, 2024, 655:1-25.

[30] Zhangmeng Liu, Philip S. Yu. Classification, Denoising, and Deinterleaving of Pulse Streams With Recurrent Neural Networks. IEEE Transactions on Aerospace and Electronic Systems, 2019, 55(44):1624-1639.

[31] Kunihiko Fukushima. Artificial Vision by Deep CNN Neocognitron. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021, 51(1):76- 90.

[32] Hongchun Qu, Jian Zheng, Xiaoming Tang. Effects of loss function and data sparsity on smooth trajectory extraction with deep model. Expert Systems With Applications, 2022, 198:1-10.

[33] Brenier, Yann. Polar factorization and monotone rearrangement of vector-valued functions. Commun. Pure Appl. Math. 1991; 44(4):375-417.

[34] Kehua Su, Na Lei, Ming Ma, et al. Focus+ Context Visualization Based on Optimal Mass Transportation. Best Paper Award, Chinagraph 2018.

[35] Xianfeng Gu, Feng Luo, Jian Sun, et al. "Variational principles for minkowski type problems, discrete optimal transport", and discrete monge-ampere equations. Asian Journal of Mathematics, 2016, 20(2):383-398.