# T-distribution Random Nearest Neighbor Embedding Combined with Entropy Weight

Yongxin Tang, Limin Zou

***Abstract*—Dimensionality reduction is the process of transforming high-dimensional data into low-dimensional representations, aiming to preserve as much important information as possible while reducing data complexity. It plays an important foundational role in data analysis and machine learning. Feature extraction generates new features by transforming raw data to preserve important information while reducing the dimensionality of the data. T-SNE is a non-linear feature extraction algorithm that is particularly suitable for visualizing high-dimensional data. The traditional t-SNE algorithm does not fully consider the importance of each feature. In view of this, this paper proposes a t-SNE dimensionality reduction method combined with entropy weight. Firstly, the entropy weight method is used to calculate and allocate the weight of each feature, and then the weighted features are subjected to t-SNE dimensionality reduction processing. Comparative experiments were conducted on five low-dimensional datasets, five medium-dimensional datasets, and four high-dimensional datasets. The experimental results show that the t-SNE algorithm combined with entropy weight outperforms the traditional t-SNE algorithm in terms of KL divergence values. Therefore, the data distribution after dimensionality reduction using the algorithm proposed in this paper is more similar to that in the original high-dimensional space.**

***Index Terms*—entropy weight, feature weight, t-SNE algorithm, KL divergence**

## I. Introduction

DATA mining is a technique that analyzes large-scale data to reveal its inherent patterns, correlations, and trends, providing valuable information and insights[1]. In the process of data mining, high-dimensional data often leads to a curse of dimensionality, making model training and prediction more difficult and time-consuming. Data dimensionality reduction is a critical stage in data mining, aimed at improving the efficiency of analysis and modeling by optimizing the structure and quality of the dataset. Effective dimensionality reduction techniques can significantly enhance the quality of data mining results and applicational effectiveness.

Data dimensionality reduction can usually be divided into feature selection and feature extraction. Feature selection reduces data dimensionality by selecting the most important subset of features in the dataset, including filtering, wrapping, and embedding methods, while feature extraction transforms the original dataset into a new feature space through mathematical transformations[11]. Common methods include principal component analysis and t-distributed stochastic neighbor embedding (t-SNE). Principal component analysis maps data to a new feature space through linear transformation, where the new features are linear combinations of the original features that are independent of each other. t-SNE, on the other hand, is a nonlinear dimensionality reduction method aimed at preserving the local structure of data points in high-dimensional space and mapping them to low-dimensional space, so that similar data points in high-dimensional space remain similar in low-dimensional space.

In 2008, Maaten et al.[9] proposed the t-SNE technique for visualizing the structure of large datasets, which maps data from high-dimensional space to two-dimensional or three-dimensional space for visualization. In recent years, many scholars have conducted extensive research on the t-SNE algorithm. Wu et al.[17] proposed an online distribution network operation mode recognition method based on real-time measurement data from smart meters. This method first preprocesses the node voltage and power data obtained from the smart meter, calculates the pairwise voltage difference between nodes, and then applies the t-distributed random nearest neighbor embedding (t-SNE) algorithm to reduce the dimensionality of the voltage difference data and power data of the historical time section. Zhong et al.[19] designed a dynamic analysis system for power equipment data monitoring based on t-SNE, aiming to solve the problems of insufficient monitoring of abnormal states in traditional power equipment and the inability to provide real-time warnings. Liu et al.[8] proposed a t-distribution random nearest neighbor embedding feature extraction and radial propagation clustering algorithm based on user voltage data for phase recognition. Jiang et al.[5] studied the distribution of spectral data in high-dimensional space and the principle of manifold learning for dimensionality reduction of high-dimensional linear data. They compared the differences in dimensionality reduction effects between t-SNE and principal component analysis methods for spectral data and used an improved K-nearest neighbor algorithm based on attribute value correlation distance for spectral classification. Wei et al.[16]proposed an extended t-SNE dimensionality reduction visualization algorithm, E-t-SNE, to address the problem that traditional t-SNE algorithms are only applicable to single type data and cannot effectively handle mixed attribute data. Yoshiaki et al.[18] proposed using quantum neural networks to parameterize t-SNE, in order to accurately reflect the characteristics of high-dimensional quantum data in low-dimensional space. Dmitry et al.[2] introduced various initialization strategies before dimensionality reduction, including PCA initialization, high learning rate, and multi-scale similarity kernels, and used a downsampling based initialization method when dealing with very large datasets. Maaten et al.[10] solved the problem of traditional dimensionality reduction techniques being

Manuscript received November 20, 2024; revised March 31, 2025.

This work was supported by the General Program of Natural Science Foundation of Chongqing(Grant No. CSTB2023NSCQ-MSX0223) and Chongqing Technology and Business University Research Initiation Fund(Grant No. 1956043).

Y. Tang is a postgraduate student at the School of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing, 400067, China. (e-mail: 2807518452@qq.com).

L. Zou is an associate professor at the School of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing, 400067, China. (Corresponding author, e-mail: 307376216@qq.com).

unable to handle non metric similarity data visualization by constructing a map set that displays complementary structures. Tang et al.[15] effectively reduced the computational complexity of t-SNE dimensionality reduction by constructing the nearest neighbor graph of low-dimensional spatial data. Shang et al.[12] proposed an adaptive weighted optimization method to address the problem of unsatisfactory dimensionality reduction in t-SNE when processing vibration signals, and introduced genetic algorithm (GA) to solve the local optimization problem of fruit fly optimization algorithm (FOA). GA-FOA was applied to RBFNN parameter selection to improve clustering performance. Du et al.[3] proposed an improved group weighted t-SNE algorithm to address the issue of insufficient ability of t-SNE algorithm to handle singular class samples in high-dimensional data dimensionality reduction, which is applied to the clustering and recognition of singular class samples in digital handwriting. Li[7] proposed a power function weighted t-SNE algorithm to solve the problem of insufficient measurement of high-dimensional sample similarity by Euclidean distance during dimensionality reduction, resulting in poor dimensionality reduction performance. Zou et al.[20] used K-means clustering combined with t-SNE to evaluate the transportation safety of imported natural gas in China.

The determination of feature weights helps to analyze the contribution of each feature to the overall system. By allocating weights reasonably, the impact of each feature can be evaluated more scientifically, leading to more rational decisions. The traditional t-SNE dimensionality reduction method does not consider the differences in feature weights, which may lead to poor dimensionality reduction results in datasets with significant feature weight variations. To address this issue, this paper proposes a t-SNE algorithm combined with entropy weight. This method first uses the entropy weight method to calculate and allocate feature weights, integrates the weight information into the data, and then uses the t-SNE algorithm to reduce the dimensionality of the weighted data. Finally, comparative experiments were conducted on 14 different types of datasets, and the results showed that compared with the traditional t-SNE algorithm, the algorithm proposed in this paper has a smaller KL divergence value. Therefore, the data distribution after dimensionality reduction using the algorithm proposed in this paper is more similar to the data distribution in the original high-dimensional space.

## II. T-DISTRIBUTION RANDOM NEAREST NEIGHBOR EMBEDDING

### A. Random neighbor embedding

Random Neighbor Embedding (SNE) is a machine learning algorithm used for nonlinear dimensionality reduction, which can effectively capture the complex manifold structure of the original data while reducing the vector dimension[4]. The algorithm first converts the distance into a probability distribution using Gaussian distribution in the original data space, and then specifies the mean squared error of the Gaussian distribution as $\frac{1}{\sqrt{2}}$ in the low-dimensional embedding space. KL divergence is used as a measure of similarity between two probability distributions, with smaller values indicating closer proximity between the two distributions. Therefore, KL divergence can be used to quantify the differences between high-dimensional and low-dimensional spatial distributions.

In high-dimensional space, the similarity between data point $x_i$ and data point $x_j$ is represented by the conditional probability density $p_{j|i}$ . The larger the value of $p_{j|i}$, the higher the similarity between $x_i$ and $x_j$:

$$p_{j|i} = \frac{\exp\left(-\left\|x_i - x_j\right\|^2 / 2\sigma_i^2\right)}{\sum\limits_{k \neq i} \exp\left(-\left\|x_i - x_k\right\|^2 / 2\sigma_i^2\right)}. \tag{1}$$

Among them, $\sigma_i$ is usually taken as the Gaussian mean squared error centered around $x_i$. In addition, set $p_{i|i} = 0$ and for $y_i$ in low-dimensions, specify the mean squared error of the Gaussian distribution as $\frac{1}{\sqrt{2}}$. The similarity between data point $x_i$ and data point $x_j$ is as follows:

$$q_{j|i} = \frac{\exp\left(-\left\|y_i - y_j\right\|^2\right)}{\sum\limits_{k \neq i} \exp\left(-\left\|y_i - y_k\right\|^2\right)}. \tag{2}$$

Similarly, set $q_{i|i} = 0$ to minimize the KL divergence between high-dimensional and low-dimensional spaces:

$$C = KL\left(P||Q\right) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}}. \tag{3}$$

Using gradient descent, calculate the gradient $\frac{\partial C}{\partial y_i}$ of the objective function with respect to the independent variable:

$$\frac{\partial C}{\partial y_i} = 2 \sum_j \left(p_{j|i} - q_{j|i} + p_{i|j} - q_{i|j}\right)\left(y_i - y_j\right). \tag{4}$$

Update data:

$$y^{(t)} = y^{(t-1)} + \eta \frac{\partial C}{\partial y} + \alpha\left(t\right)\left(y^{(t-1)} - y^{(t-2)}\right). \tag{5}$$

Loop iteration minimizes the loss function C to obtain low-dimensional data $y = \{y_1, y_2, \cdots, y_n\}$.

### B. T-distribution random nearest neighbor embedding

The random nearest neighbor embedding algorithm has two main problems: (1) The similarity between data points is asymmetric, that is, the similarity between data point $x_i$ and data point $x_j$ is not equal to the similarity between data point $x_j$ and data point $x_i$; (2) In low-dimensional space, data points are prone to "crowding problems", where data points from different clusters gather together and are difficult to distinguish clearly. To address these issues, the t-distributed random nearest neighbor embedding algorithm (t-SNE) has been proposed. This method not only ensures the similarity symmetry between data points, but also applies the t-distribution of long tail distribution in low-dimensional space to convert distance into probability distribution, thereby alleviating the "crowding problem" to some extent[9].

In high-dimensional space, the similarity between data point $x_i$ and data point $x_j$ is represented by a joint probability density $p_{ij}$. The larger the value of $p_{ij}$, the higher the similarity between $x_i$ and $x_j$:

$$p_{ij} = \frac{p_{i|j} + p_{j|i}}{2}. \tag{6}$$

Among them, $p_{j|i}$ represents the conditional probability density of data point $x_i$ and data point $x_j$:

$$p_{j|i} = \frac{\exp\left(-\left\|x_i - x_j\right\|^2 / 2\sigma_i^2\right)}{\sum_{k \neq i} \exp\left(-\left\|x_i - x_k\right\|^2 / 2\sigma_i^2\right)}, p_{ii} = 0. \quad (7)$$

In low-dimensional space, the similarity between data point $x_i$ and data point $x_j$ is represented by a joint probability density $q_{ij}$:

$$q_{ij} = \frac{\left(1 + \left\|y_i - y_j\right\|^2\right)^{-1}}{\sum_{k \neq l} \left(1 + \left\|y_k - y_l\right\|^2\right)^{-1}}. \quad (8)$$

Minimize KL divergence in high-dimensional and low-dimensional spaces:

$$C = KL\left(P||Q\right) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}. \quad (9)$$

Using gradient descent, calculate the gradient $\frac{\partial C}{\partial y_i}$ of the objective function with respect to the independent variable:

$$\frac{\partial C}{\partial y_i} = 4 \sum_j \left(p_{ij} - q_{ij}\right)\left(y_i - y_j\right). \quad (10)$$

Update data:

$$y^{(t)} = y^{(t-1)} + \eta \frac{\partial C}{\partial y} + \alpha\left(t\right)\left(y^{(t-1)} - y^{(t-2)}\right). \quad (11)$$

Loop iteration minimizes the loss function C to obtain low-dimensional data $y = \{y_1, y_2, \cdots, y_n\}$.

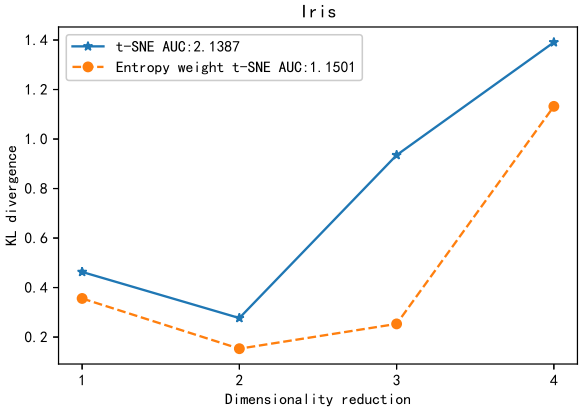


Fig. 1. Comparison of two methods on the small dataset Iris

## III. T-SNE COMBINED WITH ENTROPY WEIGHT

### A. Entropy weight method

Information entropy is the expectation of information quantity. Assuming $X$ is a random variable with $m$ values, its information entropy can be expressed as:

$$H\left(X\right) = -\sum_{i=1}^{m} p_i \log p_i. \quad (12)$$

The idea of entropy weight method is: if the information entropy of a feature is smaller, it indicates that the variation degree of the feature value is greater, so the greater the role

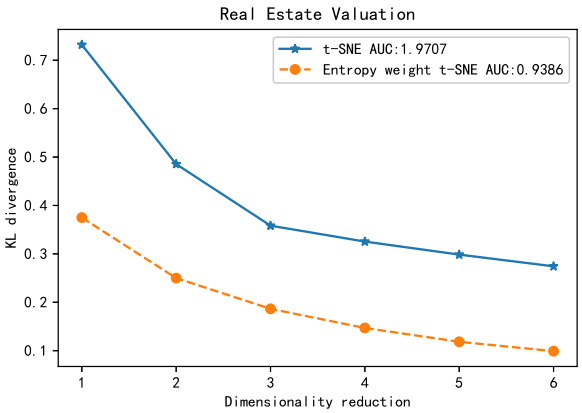


Fig. 2. Comparison of two methods on the small dataset Real Estate Valuation

of the feature, the greater the weight. The steps of entropy weight method are as follows: (i) Assuming dataset $X$ has $m$ samples, each with $n$ features:

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix}. \quad (13)$$

Among them, $x_{ij}$ is the value of the j-th feature of the i-th sample. Normalize the data:

$$y_{ij} = \frac{x_{ij} - \min(x_{ij})}{\max(x_{ij}) - \min(x_{ij})}. \quad (14)$$

(ii) Calculate the proportion of the i-th sample to a certain feature $X_j$, where $p_{ij}$ is the proportion of the i-th sample to the j-th feature:

$$p_{ij} = \frac{y_{ij}}{\sum_{i=1}^{m} y_{ij}}. \quad (15)$$

(iii) Calculate the entropy value of a certain feature$X_j$:

$$E_j = -\frac{1}{\log\left(m\right)} \sum_{i=1}^{m} p_{ij} \log p_{ij}. \quad (16)$$

(iv) Calculate the information redundancy $D_j$ of each feature:

$$D_j = 1 - E_j. \quad (17)$$

(v) Normalize the redundancy $D_j$ to obtain the weight $W_j$:

$$W_j = \frac{D_j}{\sum_{j=1}^{n} D_j}. \quad (18)$$

### B. Improved t-SNE

This paper proposes a new t-SNE dimensionality reduction method by combining entropy weight and t-SNE dimensionality reduction. The steps of this method are as follows:

(i) Assuming dataset $X$ has $m$ samples, each with $n$ features:

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{bmatrix}. \quad (19)$$

(ii) Use entropy weight method to obtain the weight of each feature:

$$W = [W_1, W_2, \cdots, W_n]^T . \tag{20}$$

(iii) Add the weight of each feature to the data of each feature to obtain a weighted dataset $Z$:

$$Z = X \times diag\,(W) . \tag{21}$$

(iv) Using t-SNE to reduce the dimensionality of weighted datasets and obtain the dimensionality reduction results.

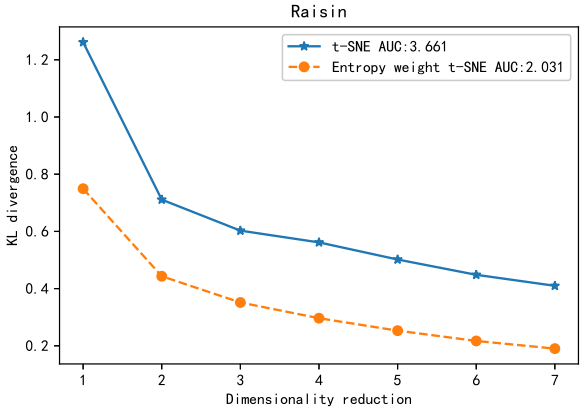


Fig. 3. Comparison of two methods on the small dataset Raisin

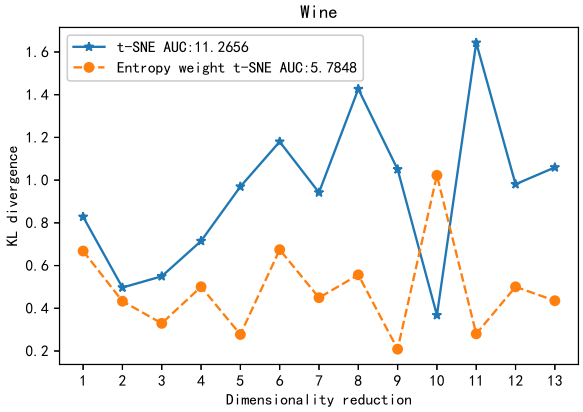


Fig. 4. Comparison of two methods on the small dataset Wine

## IV. Experimental Results and Comparative Analysis

### A. Dataset description

The dataset can be classified based on the number of features. Generally speaking, a dataset with 0 to 19 features is considered a small dataset, a dataset with 20 to 49 features is considered a medium dataset, and a dataset with 50 or more features is considered a large dataset[6][13][14]. To ensure the representativeness of the dataset, this study selected 14 datasets from UCI as research objects(https://archive.ics.uci.edu/). Table 1 shows the names, sample sizes, and number of features of these datasets.

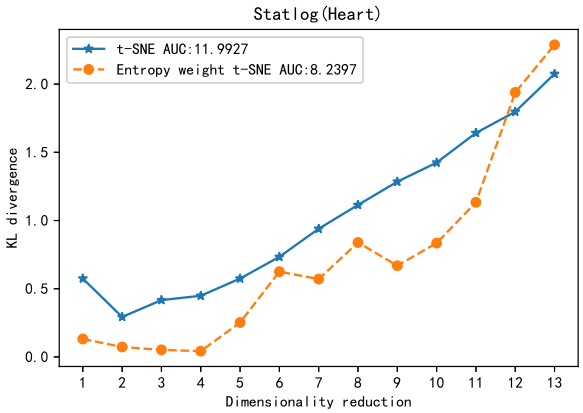


Fig. 5. Comparison of two methods on the small dataset Statlog(Heart)

TABLE 1
DATASET DESCRIPTION

| type | name | sample size | feature number |
|---|---|---|---|
| small | Iris | 150 | 4 |
| small | Real Estate Valuation | 414 | 6 |
| small | Raisin | 900 | 7 |
| small | Wine | 178 | 13 |
| small | Statlog(Heart) | 270 | 13 |
| medium | Breast cancer | 569 | 30 |
| medium | IR Temperature | 1020 | 33 |
| medium | Horse Colic | 368 | 27 |
| medium | Dermatology | 366 | 34 |
| medium | Glioma Grading Clinical | 839 | 23 |
| large | Spambase | 401 | 57 |
| large | Lung cancer | 32 | 56 |
| large | Mice Protein Expressionin | 1080 | 80 |
| large | Connectionist Bench | 208 | 60 |

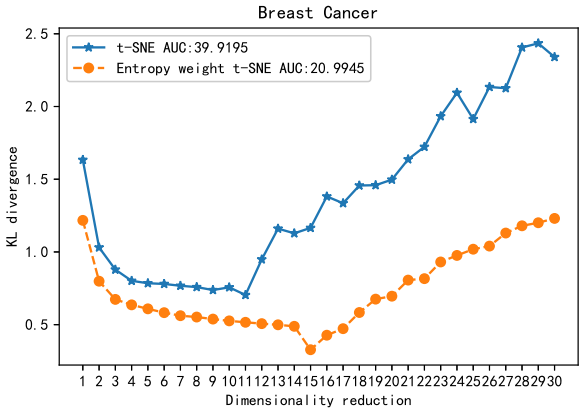


Fig. 6. Comparison of two methods on the medium dataset Breast cancer

### B. Result analysis

Using t-SNE combined with entropy weight and traditional t-SNE for dimensionality reduction on different datasets, KL divergence is used as a measurement standard to obtain the comparison values of KL divergence between the two methods in different dimensions. For ease of observation, the dimensionality reduction dimension is used as the x-axis, and the KL divergence value is used as the y-axis for visualization. Firstly, the performance of the two methods on small datasets is shown in Fig 1-Fig 5. Observation

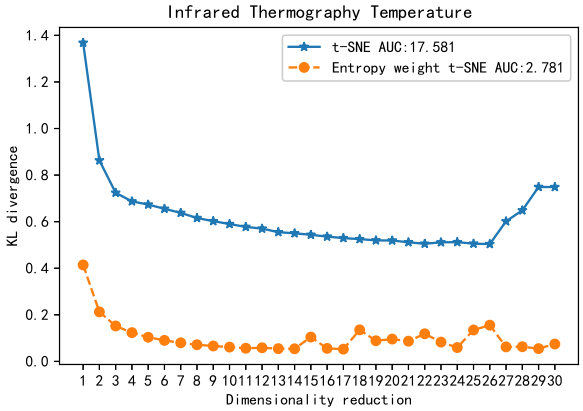


Fig. 7. Comparison of two methods on the medium dataset Infrared Thermography Temperature

shows that two curves intersect in the graph. In order to objectively compare the advantages and disadvantages of the two methods, the Area Under the Curve (AUC) under the two curves is calculated.

As illustrated in Fig 1-Fig 5, the KL divergence values for the two methods across different dimensionalities indicate that t-SNE combined with entropy weight outperforms traditional t-SNE on the Iris, Raisin and Real Estate Valuation datasets. In contrast, for the Wine dataset, when the number of dimensions is reduced to 10, the KL divergence value for the t-SNE with entropy weight is higher than that of traditional t-SNE. A further comparison of the AUCs reveals that the AUC for t-SNE with entropy weight (5.7848) is significantly lower than that of traditional t-SNE (11.2656). For the Statlog (Heart) dataset, t-SNE combined with entropy weight exhibits higher KL divergence values than traditional t-SNE when the dimensionality is reduced to 12 and 13. AUC comparisons for this dataset also show that the AUC for t-SNE with entropy weight (8.2397) is lower than that of traditional t-SNE (11.9927). These results suggest that t-SNE combined with entropy weight performs more effectively on small datasets compared to traditional t-SNE.

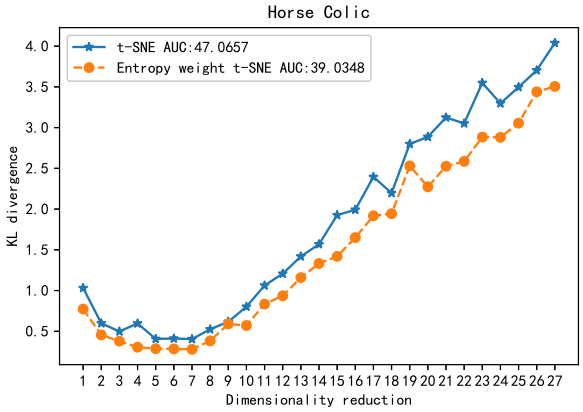


Fig. 8. Comparison of two methods on the medium dataset Horse Colic

The medium and large datasets contain a large number of features. When the number of features does not exceed 30, the KL divergence values for both methods are described

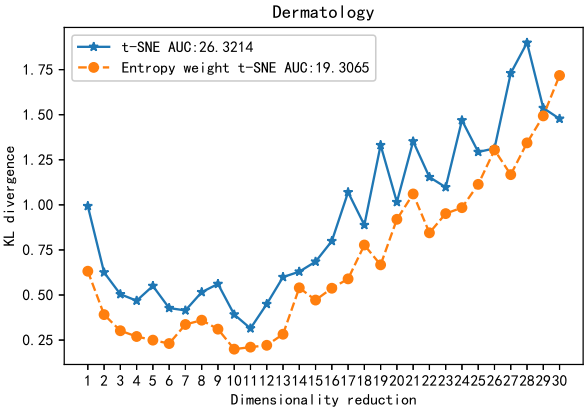


Fig. 9. Comparison of two methods on the medium dataset Dermatology

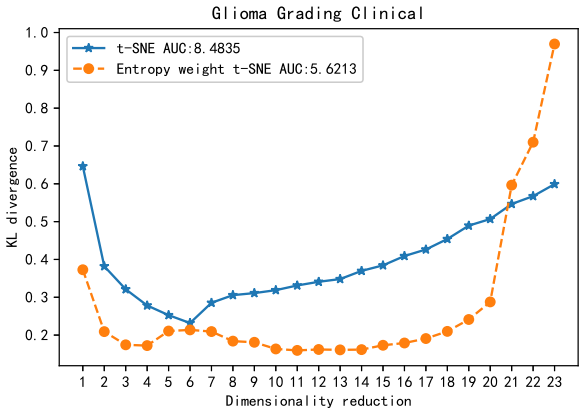


Fig. 10. Comparison of two methods on the medium dataset Glioma Grading Clinical

at each dimensionality reduction level, as shown in the graph. However, when the number of features exceeds 30, to account for computational cost, the maximum dimensionality reduction is limited to 30 for both methods. Fig 6-Fig 10 illustrates the performance of t-SNE combined with entropy weight and traditional t-SNE on a medium dataset.

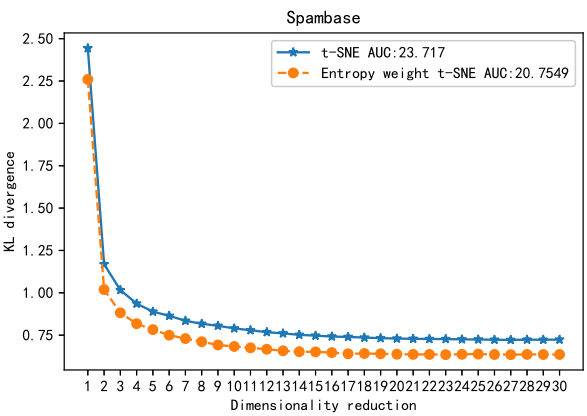


Fig. 11. Comparison of two methods on the large dataset Spambase

As illustrated in Fig 6-Fig 10, for the datasets Breast Cancer, Infrared Thermography Temperatures, and Horse Colic, t-SNE combined with entropy weight yields lower

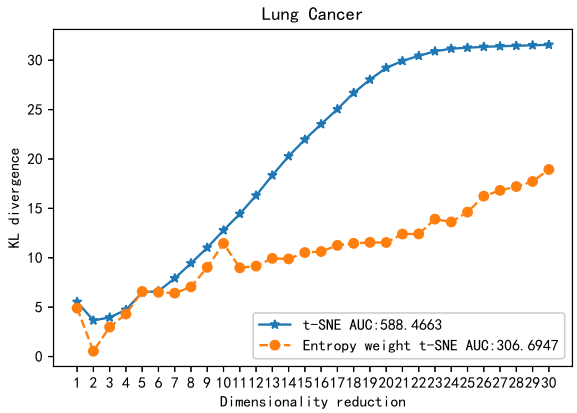


Fig. 12. Comparison of two methods on the large dataset Lung Cancer

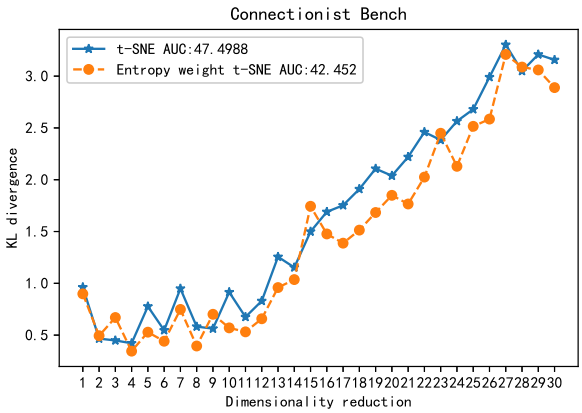


Fig. 14. Comparison of two methods on the large dataset Connectionist Bench

KL divergence values across all dimensions compared to traditional t-SNE. For the Dermatology dataset, the KL divergence value of t-SNE with entropy weight exceeds that of traditional t-SNE only when the dimensionality is reduced to 30. A further comparison of the Area Under the Curve (AUC) for both methods reveals that t-SNE combined with entropy weight achieves an AUC of 19.3065, which is lower than the AUC of traditional t-SNE (26.3214). For the Glioma Grading Clinical dataset, the KL divergence value of t-SNE with entropy weights is higher than that of traditional t-SNE only when the dimensionality is reduced to 21, 22, and 23. The AUC of this method is 5.6213, which is lower than the AUC of traditional t-SNE (8.4835). These results suggest that t-SNE combined with entropy weight demonstrates superior performance over traditional t-SNE on medium datasets.

For high-dimensional datasets with more than 49 features, this paper only considers the KL divergence values of the two methods when the dimensionality reduction does not exceed 30, as shown in Fig 11-Fig 14.

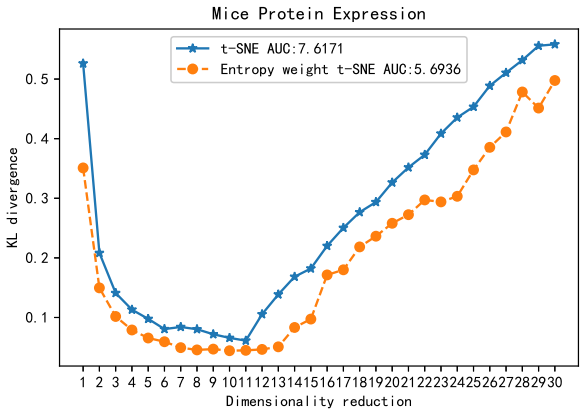


Fig. 13. Comparison of two methods on the large dataset Mice Protein Expression

Fig 11-Fig 14 demonstrates that, for the Lung Cancer, Mice Protein Expression, and Spambase datasets, t-SNE combined with entropy weight achieves lower KL divergence values across all dimensions compared to traditional t-SNE. However, for the Connectionist Benchmark dataset, t-SNE combined with entropy weight exhibits higher KL divergence values than traditional t-SNE when the dimensionality reduction is set to 3, 9, 15, and 23. Further comparison of the Area Under the Curve (AUC) for both methods reveals that t-SNE combined with entropy weight yields an AUC of 42.452, which is lower than the AUC of traditional t-SNE (47.4988). These results indicate that t-SNE combined with entropy weight outperforms traditional t-SNE on large datasets.

## V. Conclusion

This paper proposes a t-SNE dimensionality reduction algorithm that incorporates entropy weight. Comparative experiments were conducted on 14 distinct datasets, and the results demonstrated that the KL divergence values achieved by the proposed algorithm were lower, indicating superior performance compared to traditional t-SNE. However, further analysis revealed some limitations of the proposed algorithm, particularly its suboptimal performance on the image dataset MNIST, as shown in Fig 15. A possible explanation for this is that the entropy weight method may not effectively capture the feature weights in image data. In future work, we plan to explore alternative methods for calculating feature weights, with the aim of enhancing the performance of the proposed algorithm on image datasets.

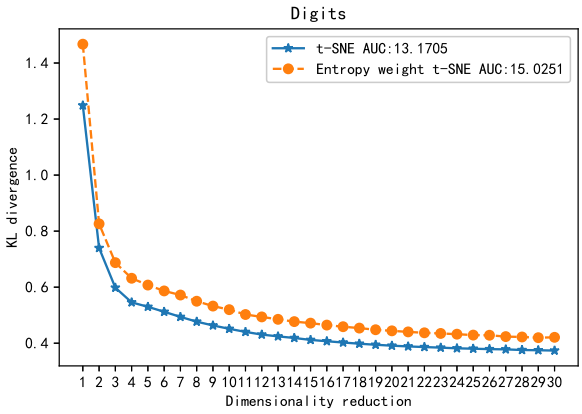


Fig. 15. Comparison of two methods on MNIST dataset

## REFERENCES


[1] P. Deng, A. Chaudhury, "An Illustration of Using Adaptive Data Mining to Develop Strategic Knowledge Bases for Student Retention", IAENG International Journal of Computer Science, vol. 50, pp. 930-940, 2023.

[2] K. Dmitry, B. Philipp, "The art of using t-SNE for single-cell transcriptomics", Nature Communications, vol. 10, pp. 5416, 2019.

[3] F. Du, B. Wang, J. Xue, et al, "Research of clustering method for handwritten numerals of oddity samples based on grouped weighted t-SNE algorithm", Journal of Chinese Computer Systems, vol. 39, pp. 2729-2734, 2018.

[4] G. Hinton, S.Roweis, "Stochastic neighbor embedding", In Proceedings of the 15th International Conference on Neural Information Processing Systems, MIT Press, Cambridge, MA, USA, vol. 1, pp. 244-268, 2018.

[5] B. Jiang, Z. Zhap, S. Wang, "Research on dimensionality reduction and classification of stellar spectra based on t-SNE", Spectroscopy and Spectral Analysis, vol. 40, pp. 2913-2917, 2020.

[6] M. Kudo, J. Sklansky, "Comparison of algorithms that select features for pattern classifiers", Pattern Recognition, vol. 33, pp. 24-41, 2000.

[7] J. Li, "Gait recognition based on power function weighted t-SNE algorithm", Harbin Engineering University, 2023.

[8] S. Liu, C. Wang, Z. Zou, et al, "Based on t-SNE dimensionality reduction and radial propagation clustering algorithm phase identification of low-voltage distribution network", Electric Power, vol. 56, pp. 108-117, 2023.

[9] L. Maaten, G. Hinton, "Visualizing data using t-SNE", Journal of Machine Learning Research, vol. 9, pp. 2579-2605, 2008.

[10] L. Maaten, G. Hinton, "Visualizing non-metric similarities in multiple maps", Machine Learning, vol. 87, pp. 33-55, 2012.

[11] T. Saw, W. Oo, "Ranking-based Feature Selection with Wrapper PSO Search in High-Dimensional Data Classification", IAENG International Journal of Computer Science, vol. 50, pp. 32-41, 2023.

[12] Q. Shang, X. Huang, D. Shen, et al, "Fault diagnosis of diesel engine based on improved t-SNE and RBFNN", Ship Engineering, vol. 45, pp. 91-97, 2023.

[13] N. Sureja, B. Chawda, A.Vasant, "An improved K-medoids clustering approach based on the crow search algorithm", Journal of Computational Mathematics and Data Science, vol. 3, pp. 100034, 2022.

[14] M. Tahir, A. Bouridane, F. Kurugollu, "Simultaneous feature selection and feature weighting using Hybrid Tabu Search/K-nearest neighbor classifier", Pattern Recognition Letters, vol. 28, pp. 438-446, 2007.

[15] J. Tang, J. Liu, M. Zhang, et al, "Visualizing large-scale and high-dimensional data", In Proceedings of the 25th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, vol. 1, pp. 287-297, 2016.

[16] S. Wei, X. Li, Y. Zhang, et al, "Dimension reduction and visualization of Mixed-Type data based on E-t-SNE", Computer Engineering and Applications, vol. 56, pp. 66-72, 2020.

[17] M. Wu, J. Wang, Y. Zhu, et al, "Online identification of active distribution network operation mode based on t-SNE dimensionality reduction and clustering", Electric Power Construction, vol. 44, pp. 52-60, 2023.

[18] K. Yoshiaki, M. Kosuke, F. Keisuke, et al, "Parametric t-stochastic neighbor embedding with quantum neural network", Physical Review Research, vol. 4, pp. 043199, 2022.

[19] J. Zhong, Y. Zhang, Y. Ren, et al, "Dynamic monitoring data of power equipment based on t-SNE algorithm analysis system", Journal of Kashi University, vol. 44, pp. 40-43, 2023.

[20] L. Zou, Y. Tang, "Prediction of China's natural gas import volume and transportation safety evaluation based on machine learning", Journal of Industrial Technological Economics, vol. 44, pp. 108-118, 2025.