Journal of Networking Technology

DLINE JOURNALS

Print ISSN: 0976-898X
Online ISSN: 0976-8998

JNT 2025: 16 (1)
https://doi.org/10.6025/jnt/2025/16/1/9-16

Analysis of the Legalization and Structure Construction of Multi-Modal Green Industry Development

Bing Gao
Zibo Vocational Institute
Zibo, Shandong, 255300, China
zbvcgaobing@163.com

ABSTRACT

The legalization and structure construction of multi-modal green industry development is a crucial issue in the current context of environmental protection and sustainable development. In this context, this paper explores how to promote the development of green industries by establishing sound legal and institutional systems. Measures such as formulating relevant regulations, strengthening environmental supervision, promoting environmental certification, enhancing talent cultivation, and advancing green procurement are proposed. By improving relevant policies and regulations and strengthening the regulation of various green industries, this approach effectively incentivizes their healthy, stable, and efficient operation, thus achieving dual sustainability in the economic, social, and environmental domains. Additionally, it contributes to enhancing China's competitiveness and influence in the global environmental protection field.

Keywords: Green Industry Development, Big Data, Multi-Modal, Sustainable Development

Received: 8 September 2024, Revised 22 October 2024, Accepted 14 November 2024

1. Introduction

As global environmental issues become increasingly severe, developing green industries has become a common concern for governments and enterprises worldwide [1]. The development of green industries can not only reduce environmental pollution and damage but also promote sustainable economic development. However, due to environmental costs and technological limitations, many enterprises find it challenging to implement ecological measures and engage in green industries voluntarily. Therefore, governments and relevant authorities must promote green industry development by establishing a sound legal and institutional system. Developing green industries requires adopting environmental technologies and measures, which often entail high investment costs and technological requirements [2]. In some areas, the market demand for green products is insufficient, leading to enterprises lacking enthusiasm for engaging in green industries. In certain regions and fields, there is

relatively low public awareness and understanding of environmental issues, which also affects the development of green industries. The development of green industries involves multiple disciplines, such as environmental science, economics, political science, etc. Hence, interdisciplinary research and collaboration are needed to address various issues in the development of green industries comprehensively. In this regard, big data multi-modal algorithms can address these challenges and contribute to the legalization and structure construction of green industry development. Multi-modal algorithms are algorithms that simultaneously process multiple types of data or features, such as sound, images, and text data. Common multi-modal algorithms include deep learning, which uses multi-layer neural networks to simulate the human brain's learning process, thus achieving automatic feature extraction and classification [3]. Ensemble learning combines multiple models to improve performance. Convolutional neural networks use convolutional layers to extract image features, enabling image classification and recognition. Cross-modal algorithms involve processing data of one modality mapped onto another modality's data, such as image data onto text data for processing. Therefore, it can be said that cross-modal algorithms belong to a subset of multi-modal algorithms. These multi-modal algorithms can simultaneously process multiple data types or features, providing more comprehensive and accurate decision support. In summary, the research background for the legalization and structure construction of multi-modal green industry development includes constraints in environmental technology and costs, uncertainties in policies and regulations, insufficient market demand, low ecological awareness and social cognition, and the necessity of interdisciplinary research and collaboration. These background factors provide direction and focus for research, contributing to the healthy development of green industries.

2. Related Work

With the development and deepening of technology applications, data acquisition and utilization are constantly changing. Multi-modal data fusion technology is becoming an increasingly important data processing method [5]. By integrating data of different types and from different sources, multi-modal data fusion can provide more comprehensive and rich information, offering more substantial support for the legalization and structure construction of green industry development. In the field of multi-modal research, previous researchers have conducted valuable studies and explorations. Ugwu et al. proposed a multi-modal biometric feature fusion method based on Independent Component Analysis (ICA). This method first preprocesses and extracts features from each modality's data, then uses the ICA algorithm to fuse the features of different modalities and recognizethem through classifiers. The results show that this method achieves high accuracy and robustness in multi-modal biometric feature recognition tasks [6]. Altay et al. presented a deep learning-based multi-modal emotion recognition method. This method uses Convolutional Neural Networks (CNN) for feature extraction and classification of speech and facial expression data and Support Vector Machines (SVM) for text data classification. The results demonstrate high accuracy and generalization performance in multi-modal emotion recognition tasks [7]. Abarajithan et al. proposed a joint learning-based multi-modal image classification method. This method uses CNN for feature extraction and classification of image data and uses joint training to combine and fuse features from different modalities. The results indicate high accuracy and robustness in multi-modal image classification tasks [8]. Lakshmi et al. introduced a multi-modal image classification method based on sparse coding. This method uses sparse coding algorithms for feature extraction and joint coding of different modalities' image data and uses classifiers for classification. Experimental results show good accuracy and generalization performance in multi-modal image classification tasks [9]. Sokhangoee and his team developed a new multi-modal image search technology based on autocorrelation matrices, which can effectively search various modalities of data, achieving more accurate search results. The results demonstrate good accuracy and recall rate in multi-

modal image retrieval tasks. These literature examples illustrate the applications and challenges of multi-moda algorithms in different fields [10]. Their research directions include multi-modal biometric feature recognition, emotion recognition, image classification, image retrieval, etc. The commonality among these methods is that they simultaneously process multiple types of data or features and fuse and classify them using different algorithms. These methods are essential in improving recognition accuracy, robustness, and generalization performance [11,12]. Furthermore, with the development and application of deep learning technology, many deep learning-based multi-modal algorithms have emerged in recent years, such as deep co-representation learning, multi-modal autoencoders, etc. These algorithms achieve applications in various fields, such as speech recognition, image classification, legalization and structure construction of green industry development, etc., by learning joint representations and similarity measures between multi-modal data [13]. In conclusion, multi-modal algorithms constitute a field full of challenges and opportunities, with a wide range of applications in computer vision, speech processing, biometric recognition, emotion computing, and many other fields. With technological progress, the application scope of multi-modal algorithms continues to expand, and their complexity and intelligence are continuously enhanced, providing more comprehensive, accurate decision support and legal guarantees for the development of green industries.

3. Multi-Modal System

3.1 Multi-modal Pretrained Models

Multi-layer transformers are widely used in many current multimodal pre-trained models (MM-PTMs). The input of each modality is first extracted into feature embeddings by independent encoders and then interacts with other modalities [14,15]. Based on the multi-modal information fusion, there are two types of MM-PTMs: single-stream and cross-stream. We will present these two architectures separately. In the single-stream architecture, multi-modal inputs, such as images and text, are processed and fused equally in a unified model. Single-modal features extracted from each modality are tokenized and concatenated to serve as inputs to the multi-modal transformer for fusion, as shown in Figure 1. In the transformer, the Multi-Head Self-Attention (MHSA) mechanism is typically used to interactively fuse single-modal features, which are then output from the transformer's class token as multi-modal fusion features.

Figure 1 presents the single-stream pre-trained multi-modal model architecture. In the cross-stream architecture, different modalities' features are extracted in parallel by independent encoders and then aligned through self-supervised contrastive learning. The pre-trained model before training obtains aligned single-modal features rather than fused multi-modal features [16]. The multi-modal fusion features are obtained by concatenating the single-modal features and inputting them into a Multi-Layer Perceptron (MLP) for pretraining target learning.

3.2 Pre-trained Model Algorithms

The design of learning objectives is a crucial step in multi-modal pre-training. The following learning objectives are currently proposed, including contrastive loss, generative loss, etc. The contrastive loss (CS) function usually constructs positive and negative training samples, which have been widely used in multi-modal tasks. For example, CLIP and ALIGN use contrastive learning loss for training. The authors of *VinVL* use a three-way contrastive loss for pre-training, replacing the binary contrastive loss used in the Oscar model. The contrastive loss is defined as shown in Formula 1.

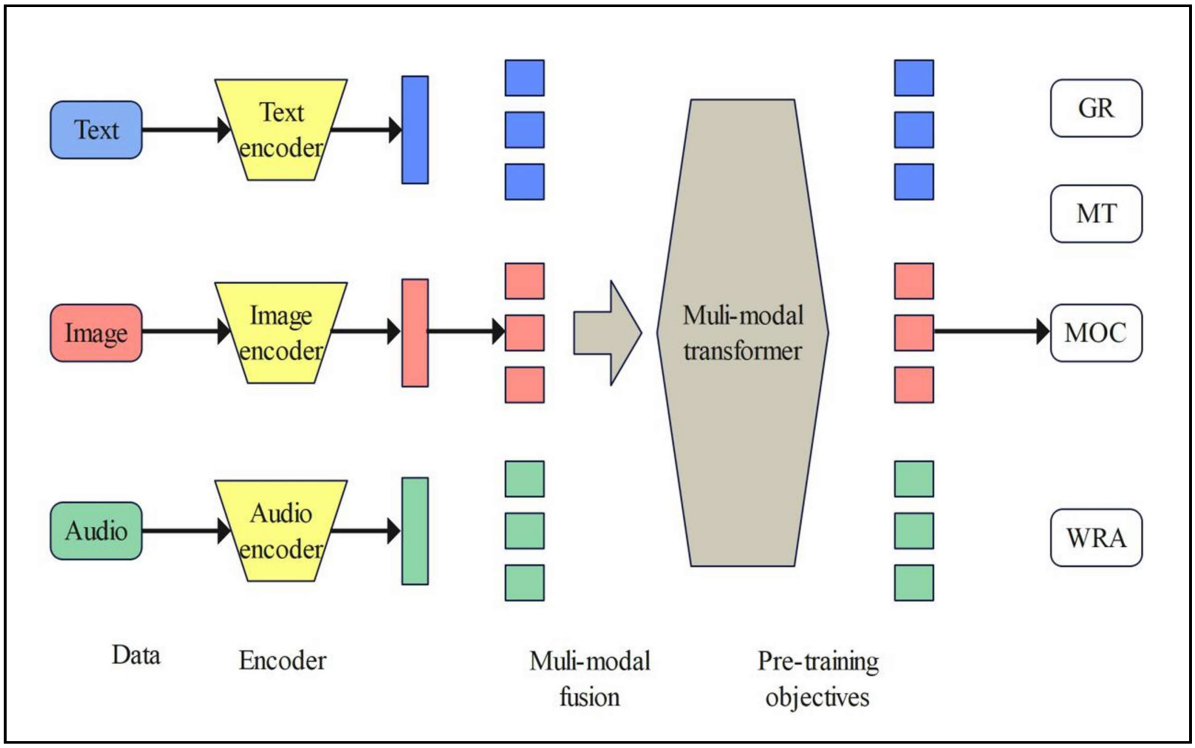

Figure 1. Architecture of Single Stream Pre-trained Multimodal Model

$$\mathcal{L}_{i2t} = -\frac{1}{N}\sum_{i}^{N}\log\frac{\exp\left(x_i^{\mathrm{T}} y_i / \sigma\right)}{\sum_{j=1}^{N}\exp\left(x_i^{\mathrm{T}} y_j / \sigma\right)} \quad (1)$$

Due to the explicit or implicit alignment relationship between different modalities, the Modality Matching Loss (MML) has been widely used in large multi-modal pre-trained models. Positive and negative image pairs are extracted, and the model is trained to predict whether the given sample pairs are aligned (predict the matching score). Unlike conventional negative image-text pairs, image-text matching is designed with hard negative images by selecting the highest TF-IDF similarity (i.e., ITM-hn).

3.3 Masked Language Framework

Masked Language Modeling (MLM) is another widely used pre-training objective, where researchers usually randomly mask and pad input words with special tokens. Surrounding words and corresponding image regions serve as references for predicting masked words. The SIMVLM is trained using Prefix Language Modeling (PrefixLM), which conducts bidirectional attention on the prefix sequence and autoregressive decomposition on the remaining tokens. These words are denoted as w = {*x1, ⋯, xK*}, and image regions are denoted as *v* = {*v1, ⋯, vT*}. For MLM, the input word *xm* is masked with a masking index m with a certain probability *p*%. The optimization objective is to predict the remaining words *x*¬*m* of the masked word based on all image regions *v*, by minimizing the negative log-likelihood value, as shown in Formula 2.

$$\mathcal{L}\big|_{MLM}(\theta) = -\mathrm{E}_{(x,v)}\log P_\theta\left(x_m \mid x_{\neg m}, v\right) \quad (2)$$

3.4 Masked Object Classification

Masked Object Classification (MOC) mainly masks visual images with zero values. Then, the labels predicted by the object detector are often taken as the ground truth labels. This pre-training objective is widely used, similar to MLM, where image regions can be masked by masking their visual features with $p\%$ probability. The goal is to predict the object category im of the masked image region v. The encoder output im of the masked image region v is fed into a fully connected (FC) layer to predict scores for T object classes, which are further transformed into a normalized distribution $g\theta(v)$ through a *softmax* function. The final objective for im is as shown in Formula 3. Additionally, the Action Prediction (AP) objective is to evaluate whether the agent developed for Visual Language Navigation (VLN) can accurately select the correct action based on the current image and instruction. XGPT uses an Image-Conditioned Denoising Autoencoder (IDA) with attention matrices to align bottom-level image-text pairs. Even without the previous length of masked fragments, IDA can reconstruct the entire sentence successfully. The Attribute Prediction (AttP) is used to recover the masked markers of attribute pairs. ERNIEViL uses Relationship Prediction (RelP) to predict the probability of recovering masked relationship tokens for each masking relationship token.

$$\mathcal{L}_{MOC}(\theta) = -\mathrm{E}_{(w,v)} \sum_{i=1}^{M} CE\left(c\left(v_m^i\right), g_\theta\left(v_m^i\right)\right) \tag{3}$$

4. Experimental Design and Analysis

4.1 Experimental Design

This study aims to comprehensively evaluate the production, consumption, and environmental aspects of the green industry and its related indicator system to delve into the sustainability of green development. Through comparative analysis, we found that a comprehensive indicator system can better reflect the actual situation and demonstrate its overall characteristics. To ensure its effectiveness, we should select indicators based on five basic principles: scientific, comprehensive, practical, flexible, systematic, and autonomous. Based on relevant research and information, we have formulated a comprehensive green development catalog, which includes three main parts: ecological construction, economy, and human social technological progress. To this end, we have developed a set of 19 different, carefully designed, and realistic evaluation indicators. Through rigorous review and testing, we found significant deviations in each unit's indicators, which can better identify and measure different evaluation subjects, thus making more objective choices and ensuring the accuracy of measurements. At the same time, the evaluation indicators were assessed in our model. In today's society, we must recognize that innovation is crucial for achieving green development. Therefore, we must combine the two to make our green industry more dynamic and always adhere to innovation-driven and sustainable development principles, thus achieving true green development. To better evaluate the development of the green industry, we divide it into three basic levels: production, consumption, and environment. Among them, we also emphasize some indicators, such as the number of patents granted, which reflect the level of industrial development and aim to promote the healthy development of the industry.

4.2 Experimental Analysis

After calculations, the standard deviation of the energy consumption elasticity coefficient is only 0.269, indicating that this value is relatively low among all indicators, suggesting that the discriminatory power of this indicator is weak. Even though it has significant theoretical significance, its influence on ranking is not substantial for several units. Therefore, it can be discarded to construct a green industry evaluation indicator system consisting of 18

indicators to reflect the actual situation better. By comparing the loss functions in Figure 2, we can conclude. After running the evaluation indicators in the model, an error analysis of the modeling results was also conducted.

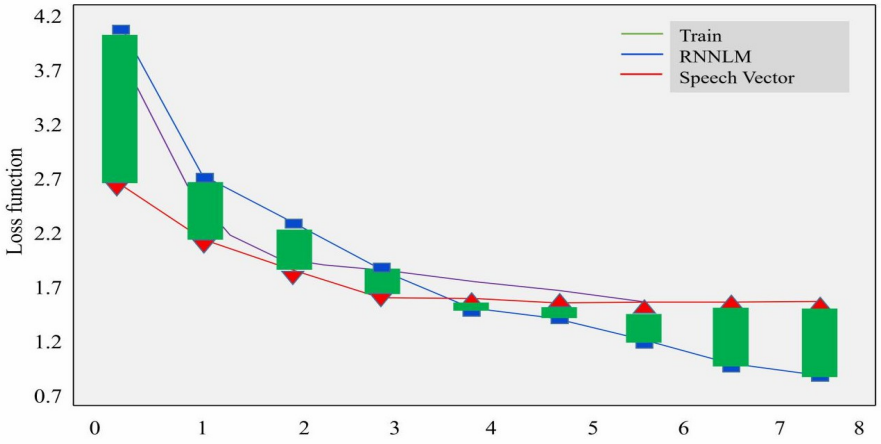

Figure 2. Comparison of Loss Functions

The effectiveness of the heterogeneous network model transfer method was tested by comparing it with the results of the subsequent two experiments. In the eight experiments, when the training loss becomes small and stabilizes, the model has essentially converged after approximately 3000-4000 iterations. In this experiment, we utilized the L1-SOFT-P method to optimize the multi-modal model transfer shear rate. We used the *softmax* layer as the feature constraint layer and parameterized L1 to train the dual-stream CNN. We only transferred the features of the target domain branch model to achieve the best model transfer effect. We applied model pruning operations before transferring the target domain branch model to the target domain. The model was pruned according to different pruning rates, and its relevant parameters were adjusted to obtain the final experimental results. Due to the large number of rows in Figure 2, we used two different styles of short dashed lines and long dashed lines for better expression of the experimental results. The short dashed lines indicate that the transfer effect significantly decreased after pruning, while the long dashed lines indicate that the effect remained unchanged. This was done to facilitate a better understanding of the experimental results. As shown in Figure 3, the transfer effect was significantly improved when using solid line pruning technology. The model pruning operation increased the model transfer effect by approximately three percentage points.

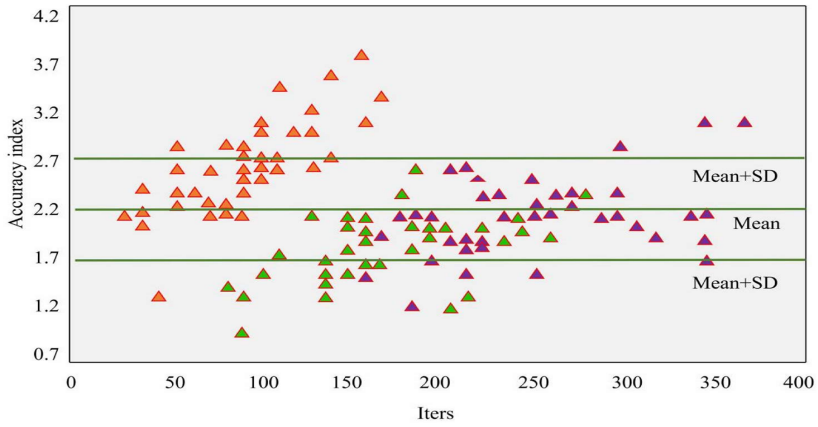

Figure 3. Experimental results of model shear rate

5. Conclusions

Through a detailed study of the 19 indicators, we have created a new evaluation indicator system that accurately measures the level of green industry development. Its unique identification capability provides us with a complete and up-to-date evaluation standard, enabling us to assess and guide the development of the green industry effectively. This system includes three main parts: production, consumption, and environment. The evaluation indicator system embodies the concept of ecological civilization and green development in current society and helps us formulate better and implement policies related to the green industry. By adjusting the values of each indicator accordingly, we can more intuitively represent the differences between them and better interpret the significance of these indicators. We also compare the weights after two adjustments to demonstrate the effectiveness of this system. Through vivid descriptions, the development trends of different indicators are clearly shown. This provides excellent help and advantages for legalizing and institutionalizing green industry development.

References

[1] Geng, Y. J. L. (2021). EARC: Evidential association rule-based classification. *Information Sciences: An International Journal, 547(1).*

[2] Basysyar, F. M., Dwilestari, G., Bahtiar, A., et al. (2021). Market basketball analysis algorithm for determining product association. *IOP Conference Series: Materials Science and Engineering, 1088(1),* 012040 (6pp).

[3] Shabtay, R. D. I. (2021). A guided FP-Growth algorithm for mining multitude-targeted item-sets and class association rules in imbalanced data. *Information Sciences: An International Journal, 553(1).*

[4] Shao, L. (2021). Research on sports training decision support system based on improved association rules algorithm. *Security and Communication Networks, 2021(4),* 1-6.

[5] Redchuk, A. (2021). Numerical association rule mining from a defined schema using the VMO algorithm. *Applied Sciences, 11.*

[6] Ugwu, N. V., Udanor, C. N. (2021). Achieving effective customer relationship using frequent pattern-growth algorithm association rule learning technique. *Nigerian Journal of Technology, 40(2),* 329-339.

[7] Altay, E. V., Alatas, B. (2021). Differential evolution and sine cosine algorithm-based novel hybrid multi-objective approaches for numerical association rule mining. *Information Sciences, 554,* 198-221.

[8] Abarajithan, S., Mohan, S. V. (2021). Cockroach swarm optimization algorithm for high utility association rule mining. *International Journal of Swarm Intelligence Research (IJSIR), 12.*

[9] Lakshmi, N., Krishnamurthy, M. (2022). Association rule mining-based fuzzy manta ray foraging optimization algorithm for frequent itemset generation from social media. *Concurrency and Computation: Practice and Experience, 34(10),* e6790.

[10] Sokhangoee, Z. F., & Rezapour, A. (2022). A novel approach for spam detection based on association rule mining and genetic algorithm. *Computers & Electrical Engineering, 97,* 107655.

[11] Jacenków, G., O'Neil, A. Q., Tsaftaris, S. A. (2022). Indication as prior knowledge for multimodal disease classification in chest radiographs with transformers. *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI),* IEEE, 1-5.

[12] Ye, J., Zhou, J., Tian, J., et al. (2022). Sentiment-aware multimodal pre-training for multimodal sentiment analysis. *Knowledge-Based Systems, 258,* 110021.

[13] Zhao, J., Li, R., Jin, Q., et al. (2022). Memobert: Pre-training model with prompt-based learning for multimodal emotion recognition. *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* IEEE, 4703-4707.

[14] Huang, Y., Lv, T., Cui, L., et al. (2022). Layoutlmv3: Pre-training for document AI with unified text and image masking. *Proceedings of the 30th ACM International Conference on Multimedia,* 4083-4091.

[15] Feng, Z., Tang, J., Liu, J., et al. (2021). Alpha at SemEval-2021 task 6: Transformer-based propaganda classification. *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021),* 99-104.

[16] Zhou, H., Ma, T., Rong, H., et al. (2022). MDMN: Multi-task and domain adaptation-based multi-modal network for early rumor detection. *Expert Systems with Applications, 195,* 116517.