

Article

Multi-Objective Prediction of Integrated Energy System Using Generative Tractive Network

Zhiyuan Zhang and Zhanshan Wang *

The College of Information Science and Engineering, Northeastern University, Shenyang 110819, China; zzy20210401@163.com

* Correspondence: wangzhanshan@ise.neu.edu.cn

Abstract: Accurate load forecasting can bring economic benefits and scheduling optimization. The complexity and uncertainty arising from the coupling of different energy sources in integrated energy systems pose challenges for simultaneously predicting multiple target load sequences. Existing data-driven methods for load forecasting in integrated energy systems use multi-task learning to address these challenges. When determining the input data for multi-task learning, existing research primarily relies on data correlation analysis and considers the influence of external environmental factors in terms of feature engineering. However, such feature engineering methods lack the utilization of the characteristics of multi-target sequences. In leveraging the characteristics of multi-target sequences, language generation models trained on textual logic structures and other sequence features can generate synthetic data that can even be applied to self-training to improve model performance. This provides an idea for feature engineering in data-driven time-series forecasting models. However, because time-series data are different from textual data, existing transformer-based language generation models cannot be directly applied to generating time-series data. In order to consider the characteristics of multi-target load sequences in integrated energy system load forecasting, this paper proposed a generative tractive network (GTN) model. By selectively utilizing appropriate autoregressive feature data for temporal data, this model facilitates feature mining from time-series data. This model is capable of analyzing temporal data variations, generating novel synthetic time-series data that align with the intrinsic temporal patterns of the original sequences. Moreover, the model can generate synthetic samples that closely mimic the variations in the original time series. Subsequently, through the integration of the GTN and autoregressive feature data, various prediction models are employed in case studies to affirm the effectiveness of the proposed methodology.

Keywords: generative tractive network; integrated energy system; feature mining**MSC:** 37M10

Citation: Zhang, Z.; Wang, Z. Multi-Objective Prediction of Integrated Energy System Using Generative Tractive Network. *Mathematics* **2023**, *11*, 4350. <https://doi.org/10.3390/math11204350>

Academic Editor: Stavros D. Kaminaris

Received: 28 September 2023

Revised: 15 October 2023

Accepted: 18 October 2023

Published: 19 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the improvement in the efficiency of sustainable energy utilization in integrated energy systems, research related to integrated energy systems has become a trending topic. Consequently, load forecasting in integrated energy systems has also become a focal point of research. For energy systems, accurate load forecasting equals better economic benefits and more convenient optimization scheduling. However, load forecasting in integrated energy systems is not limited to multiple energy objectives. It also involves coupling between these objectives due to the interconversion of multiple energy sources within the system. From Figure 1, it can be seen that the energy conversion in the integrated energy system is very complex. This coupling, combined with the inherent random fluctuations in load, makes accurate multi-objective forecasting more challenging [1].

Multi-task learning provides a framework for addressing the multi-energy load time series in integrated energy systems, allowing for learning different energy coupling relationships through information exchange. To enhance the accuracy of multi-task learning,

various statistical methods [2,3] and machine learning techniques such as feature selection, training methods, feature extraction, attention mechanisms, data augmentation, and others [4–8] can still be effective. However, for multi-energy load time series in integrated energy systems, their sequence characteristics are also essential information. The above methods do not improve model performance based on sequence characteristics. Analyzing the sequence characteristics of multi-objective time series means that the model can generate new time series from the original time series. This new time series must be related to the original time series to have an impact on the prediction model and improve its performance. Therefore, the key question is how to use a generation model to learn the characteristics of multi-target time series to achieve the desired effect.

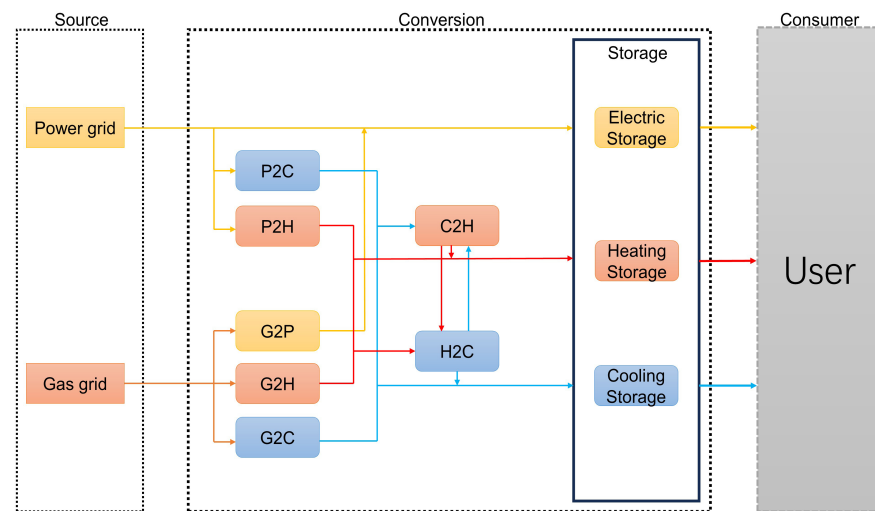


Figure 1. Conversion of multi-energy source in integrated energy system.

In recent years, language generation models have made significant progress in analyzing sequence structures. Language models are capable of generating lengthy sentences by taking lengthy sentences as input, which includes both multi-task learning and leveraging information between long sentence inputs. For textual sequence data, existing language generation models based on the transformer architecture [6] have achieved significant communicative capabilities. This demonstrates that generative models can be used to analyze structured sequence data and generate new sequence data. Moreover, in the field of natural language processing, the training process of language models has shown that the data generated by the language model itself can be used to further train the model, improving its performance. Hence, it is plausible that time-series data generated by generative models can also be employed to enhance the performance of time-series forecasting models. However, unlike text sequence data, which often have large-scale training data, time-series data are typically of smaller scale. Furthermore, it is crucial to note that time-series data often represent the outcomes of complex systems evolving over time. This is quite different from text sequence data, which typically exhibit strong logical structures like word connections. Unlike language generation models that can generate new content, time-series generation models face the challenge of lacking training labels due to the difficulty in analyzing the structure of complex systems.

In the field of computer vision, existing research has cleverly used the discrimination between real and fake data to address the issue of training generation models without sufficient labels. Generative adversarial networks (GANs) [9] rely on adversarial learning between two networks. In this framework, the generator model is trained to produce samples that closely resemble real data samples, usually from a random standard distribution. However, GANs have primarily found applications in image generation and classification tasks and are not well suited for continuous sequence data regression problems. The main goal of this paper is to design a generative model based on the principles of GANs that can

analyze the characteristics of multi-target time series. This approach generates artificial synthetic data that can be used to enhance the performance of load forecasting models in integrated energy systems.

The primary contribution of this paper lies in proposing a generative tractive network (GTN) model, comprising both a generative model and a tractive discriminative model, for feature mining on multi-target time-series data. The model leverages the autoregressive feature samples constructed based on time-series characteristics to capture specific time-series patterns. These autoregressive feature samples are then applied to the generative model to generate artificial synthetic data that reflect the analysis of the original time-series data. The process of the GTN can be seen as a process of obtaining an inverse mapping, where a fixed mapping model is used to train the generative model. This process is referred to as tractive learning. Using the above artificially synthesized time-series data results in an improvement in the performance of integrated energy system load forecasting models and enhances the robustness of the forecasting models.

2. Related Work

Integrated energy systems have strengthened their utilization of renewable energy while also introducing uncertainty and complexity [10,11]. Concerning load forecasting in integrated energy systems, various approaches have been employed to enhance prediction models, and the key is how to leverage the interaction of information between multiple energy sources while enhancing multi-task learning efficiency. For instance, in the study by [12], a K-means algorithm is utilized to extract user characteristics within an integrated energy system, followed by the application of deep belief networks for load forecasting. In [13], a hybrid model combining CNN and GRU is employed for feature extraction, and ensemble learning techniques are applied to improve multi-task learning prediction accuracy. Reference [14] leverages historical data, weather data, date data, and socioeconomic data as input features. Bootstrap resampling is employed along with machine learning methods for multi-task learning. In [15], the maximal information coefficient (MIC) is used to analyze coupling relationships between different components of integrated energy systems, which determine the inputs for the relevant model. Additionally, weather data features are incorporated, and the dataset is partitioned by season to enhance multi-task learning accuracy. Reference [16] employs Pearson coefficients to analyze the coupling relationships between various components of an integrated energy system. They utilize a parallel architecture involving CNN and GRU, allowing different features to be input from both ends and eventually combined through a fully connected layer to accomplish multi-objective forecasting tasks. Similarly, in [17], which also employs a parallel architecture with CNN and sequential networks, batch normalization layers are integrated to create a deeper prediction model for improving performance. Reference [18] adopts a multi-task learning approach alongside support vector machines to simultaneously analyze the coupling relationships among multiple energy sources. In [19], a combination of CNN and BiGRU, coupled with attention mechanisms, serves as a multi-objective prediction model. Multi-task loss functions are then optimized to achieve superior prediction performance. Reference [20] constructs multiple data compositions using various features derived from time series, statistics, and temporal changes. Weak learning models are developed for feature extraction, and these models are eventually integrated into the prediction model, leading to enhanced performance. Lastly, in [21], a wavelet transform is applied to decompose time-series data, and different prediction models are employed for high-frequency and low-frequency components. The results are then aggregated to produce multi-source load forecasts.

Regarding load forecasting in integrated energy systems, most of the cited literature employs approaches that analyze the data correlations specific to certain time series or use a combination of basic models to create complex models in order to enhance prediction model performance. For example, these approaches involve calculating data correlations [15,16], historical data correlations [14], time-series pattern correlations [3], environmental corre-

lations [15,16], date correlations [15], algorithmic feature extraction [8,12], and complex modeling techniques [17].

In summary, the existing research on multi-objective load forecasting in integrated energy systems combines various methods such as time-series analysis, analysis of integrated energy system characteristics, and machine learning methods. However, due to the inherent complexity of integrated energy systems [10,11], it is challenging to precisely analyze the entire system from within. While current research employs various correlation-based feature extraction methods, these methods may not fully capture the intricacies of complex system behaviors [22]. Furthermore, increasing the complexity of prediction models both increases computational costs and cannot guarantee the improvement of model performance on small-scale time-series data. Moreover, complex models need more hyperparameter tuning and optimization challenges. Particularly in the optimization of hyperparameters, small-scale time-series data can lead to issues of underfitting or overfitting in complex models, resulting in poor generalization accuracy of prediction models.

The existing different prediction models use data in the methods illustrated in Figure 2. Statistical models combine historical data and fluctuations for forecasting, while machine learning models utilize features with high correlations for feature extraction to complete predictions. This is all due to the difficulty in analyzing and describing the results of complex systems, where it is challenging to find a clear mathematical relationship between independent variables and predictions. In integrated energy systems, multiple energy sources such as electricity, heat, cold, and gas not only undergo mutual conversion but also exhibit their own fluctuations over time. It is challenging to establish coupling relationships for multi-objective analysis or mathematical descriptions that can be used for predictive modeling in complex systems. Even in integrated energy systems, where energy load can be viewed as a result of human activities, this merely transforms one complex variable into another, providing little help for prediction. However, attempting to analyze the characteristics of multi-energy target sequences from the perspective of a generative model, as shown in Figure 2, to obtain a new time series that can be mapped to the original multi-target time series, has yielded a desired method’s effect.

Recent literature [23] has shown that artificial data generation can enhance model performance. By constructing specific time-series data, the proposed GTN can achieve the above objectives. The advantage of the GTN lies in the fact that they do not require the design of complex network structures, time-consuming hyperparameter tuning, or the search for additional high-correlation data to increase input features. Instead, they directly analyze multi-energy target time sequences to generate artificially synthesized data that can improve predictive model performance.

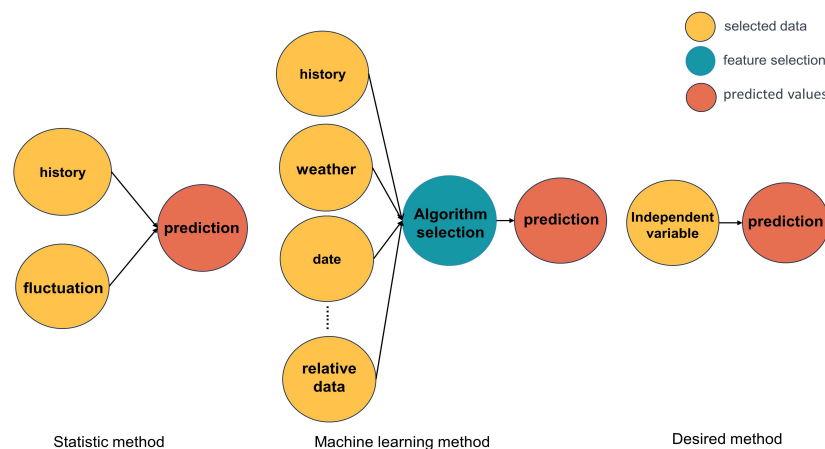


Figure 2. Description of the characteristics of different methods.

3. Analysis of Multi-Objective Time Series in Integrated Energy System

In order to ensure the specificity of the proposed GTN for multi-objective time series within integrated energy systems, specific steps are taken in both data preprocessing and the training data composition of the GTN. In data preprocessing, the amplitude differences of various energy sources are harnessed. During the training phase of the GTN, data with the capability to recognize temporal data patterns is employed. The detailed procedure is as follows.

3.1. Data Source

The data for this study were obtained from [24], consisting of original cold, heat, and electricity load data sampled at an hourly frequency for the year 2022. The dataset comprises a total of 8760 time points. However, for machine learning models, including neural networks, this sample size is relatively small. Regarding small datasets with a sample size smaller than 10,000, a split of 70% for training and 30% for validation was employed [25], while the testing set comprises the initial two weeks of the year 2023. Any anomalies present in the original load data due to sensor malfunctions or unforeseen circumstances have been treated through interpolation. The load is standardized in kilowatt (KW) units, and for precise unit conversions, refer to the explanations provided in [24].

3.2. Data Preprocessing

For machine learning, integrated energy system load prediction is evidently a multi-task learning problem, where the model needs to simultaneously learn from multiple historical datasets and predict the load values of multiple energy sources at a specific time point. Existing research in multi-task learning often involves normalizing different energy source data separately [15,16]. If the amplitude and trend of load variations for different energy sources are similar, separate normalization of different energy source data can aid in the learning process of predictive models. However, if there are substantial differences in the amplitude and trend of load variations among different energy sources, separate normalization may constrain the learning capacity of the model.

As depicted in Figure 3, within the multi-energy load data employed in this paper, the amplitude of heat load differs significantly from that of electricity and cooling load. Moreover, throughout most of the year, the amplitude variations of the heat load remain relatively consistent compared to other loads. This situation results in the heat load being of relatively secondary importance in multi-task learning predictive models. Furthermore, from the data visualization, it is evident that there are abrupt variations in the amplitude of the heat load during specific periods in 2022. If separate normalizations are applied to different loads, the abrupt variations in the heat load trend are more significant compared to the gradual variations observed in electricity and cooling loads. This could potentially affect the learning process of multi-tasking.

Therefore, after converting the load units into a unified standard, the data normalization approach adopted in this study takes into account the distinct electricity, cooling, and heating loads. The specific procedure is as follows:

$$x = \frac{x - x_{\min(\text{heat})}}{x_{\max(\text{cool})} - x_{\min(\text{heat})}} \quad (1)$$

$$x = \frac{x - x_{\min(\text{heat})}}{x_{\max(\text{cool})} - x_{\min(\text{heat})}} \quad (2)$$

where $x_{\max(\text{cool})}$ represents the maximum value among the three types of loads (electricity, cooling, and heating), and $x_{\min(\text{heat})}$ represents the minimum value among these three types of loads. This normalization approach serves the dual purpose of both reflecting the dominant role of data with larger amplitudes in the aforementioned multi-task learning and mitigating the influence of the significant fluctuations in heat load on multi-task learning within this dataset. In Figure 4, the overall trend obtained from this normalization is shown.

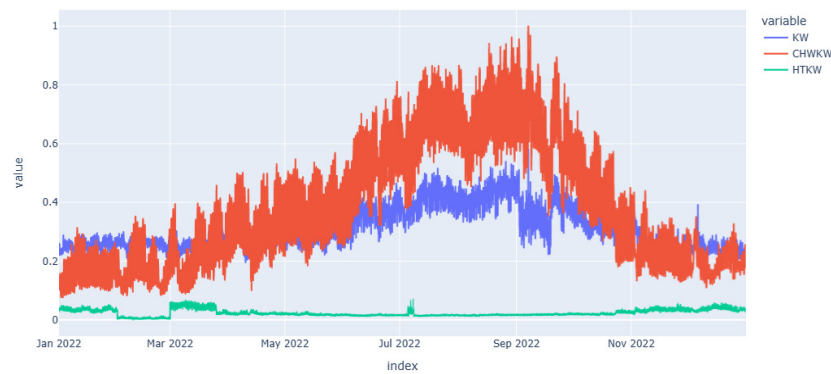


Figure 3. Normalized multi-load curves for the year 2022. In the image, “KW” represents electrical energy, “CHWKW” represents cooling energy, and “HTKW” represents heating energy. All three of them are measured in kilowatts (kW).

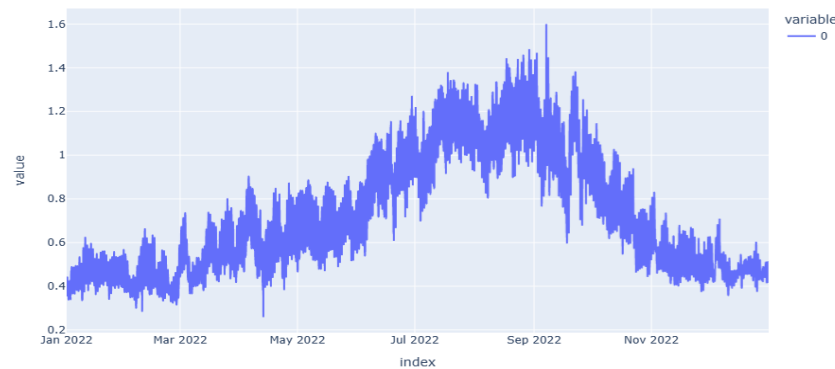


Figure 4. Total load curve for the year 2022.

3.3. Autoregressive Feature Data

For time-series prediction, historical data can significantly enhance predictive model performance. In [14], only the load value at the previous time step and other simultaneously related features are utilized for predicting the subsequent time step. In [13], feature extraction is performed using the load data from the previous four hours, combined with data from the current time point for prediction. In [16], different loads and a sequence of 24 time steps are employed as a two-dimensional feature vector, which is subjected to feature extraction using convolutional neural networks. In [15], the utilization of historical data is carried out as follows:

$$[x_{t-8760}, x_{t-168}, x_{t-24}, x_{t-2}, x_{t-1}] \rightarrow x_t \tag{3}$$

The aforementioned studies indicate that existing research primarily focuses on utilizing historical data from related time steps for predicting the target. However, the primary objective of this paper is to analyze current time-series data and generate new time-series data. In classical statistics, the autoregressive (AR) component of the ARIMA model [26] relies on historical sampled values at certain time intervals for future estimation, reflecting time-series variations through continuous time steps of historical data. Thus, to ensure that data features capture time-series characteristics, constructing autoregressive feature data can involve selecting fixed time steps. The specifics of an individual sample can be described as follows:

$$[x_{t-n}, x_{t-(n-1)}, \dots, x_{t-2}, x_{t-1}] \rightarrow y_{t-1} \tag{4}$$

Here, x_{t-1} represents the load sampled value at time point $t - 1$, n denotes the chosen time step length, and y_{t-1} signifies the newly generated data at time point $t - 1$. The autoregressive feature sample can be conceived as generating new data related to such

temporal variations for the current time point by incorporating historical sampled values from n time steps, including the current time step. Through this approach, the original time series can generate new time series with a length smaller than the original by $n - 1$.

For similar reasons as discussed above, despite not satisfying the independent and identically distributed (i.i.d) condition, autoregressive feature samples can be directly employed for predictive modeling in machine learning data. A single sample used for prediction can be described as follows:

$$[x_{t-n}, x_{t-(n-1)}, \dots, x_{t-2}, x_{t-1}] \rightarrow x_t \tag{5}$$

where the data on the left-hand side remain consistent as described earlier, with x_t denoting the time-series value at the next time step.

By transforming time series within a certain time step into data samples, this emphasis on how sequences change within a specific range will be applied to the subsequent GTN and predictive model in this paper.

4. Feature Mining Based on GTN

The generative adversarial network (GAN) [9] leverages concepts from game theory, employing a generative model and a discriminative model in a competitive process of distinguishing between real and fake instances. Through iterative optimization, the generative and discriminative models enhance each other’s performance, ultimately enabling the generative model to produce samples surpassing the discriminative model’s capabilities. This can be understood as fostering a “collaboration” between the two models during the overall training process. For specific time series, this paper draws upon this concept and employs autoregressive feature data to construct a generative tractive network (GTN) by incorporating a generative model (netG) and a discriminative model (netD) for time series feature extraction, as depicted in Figure 5. The generative model generates new artificial data while the discriminative model assesses whether the new data can replicate the original samples. The training of the generative model is guided by the feedback from the discriminative model’s outcomes. This model is referred to as the generative tractive network (GTN).

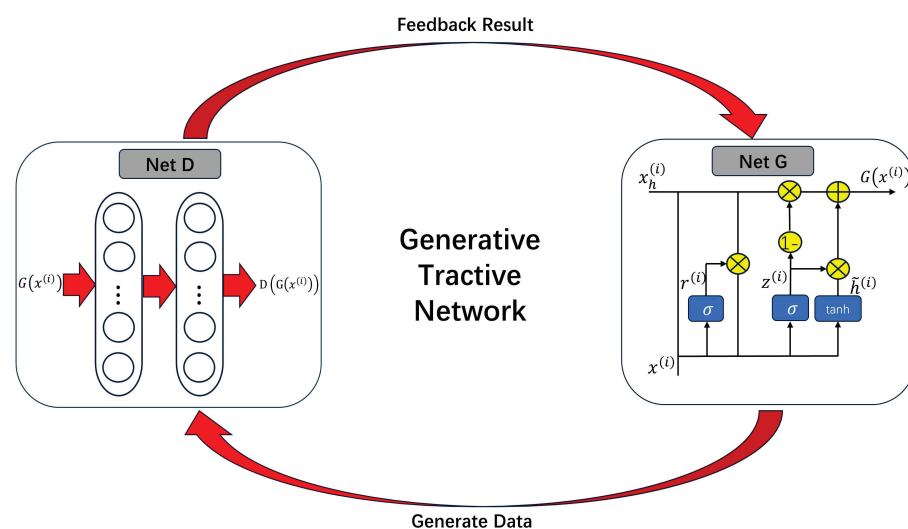


Figure 5. Generative Tractive Network.

4.1. Principle of the Generative Tractive Network

In the context of a generative tractive network, the term “tractive” can be understood as follows: when dealing with time-series data, the goal is to use a generative model to produce new time-series features that can map to the original time series. However, a generative model lacks its own labels and cannot be trained solely on its own. Therefore, in such cases, another network with labels can be used. This additional network updates itself

while simultaneously assisting in the training of the generative model. This process, where a labeled model is used to train the generative model and steer it in the desired direction, is what is meant by “tractive”.

The distinctions between tractive learning and adversarial learning within the realm of generative models are illustrated in Figure 6. In adversarial learning, the generative model generates fake samples from a random standard distribution, and the role of the discriminative model is to distinguish between real samples (labeled as 1) and the fake samples produced by the generative model (labeled as 0). The discriminative model’s classification task drives the training of the entire adversarial network. In an ideal scenario, a trained discriminative model is unable to differentiate between real samples and fake samples generated by the trained generative model, resulting in the discriminative model’s recognition probability for a mixed real–fake sample being approximately $\frac{1}{2}$, akin to random guessing.

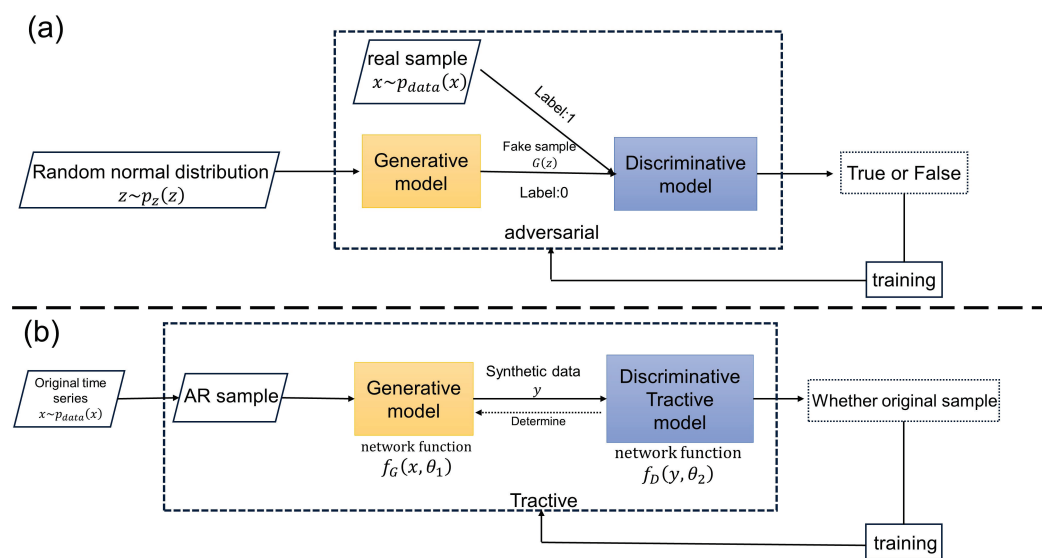


Figure 6. Difference between adversarial and tractive: (a) adversarial learning (b) tractive learning.

In tractive learning, the generative model’s objective is to analyze the original data and generate new data y highly correlated with the original data. On the other hand, the tractive discriminative model’s role is to determine whether this new data y can generate the original analyzed sample. Assuming θ represents the model parameters, ideally, the generative model function $f_G(x, \theta_1)$ and the tractive discriminative model function $f_D(y, \theta_2)$ should exhibit an inverse relationship. The newly generated data y by the trained generative model should be transformable back to the original samples, implying that this new data y are deemed highly related to the original samples.

Throughout the training process, the labeled tractive discriminative model determines the unlabeled generative model, facilitating the iterative updating of the entire network’s parameters.

The role of GAN lies in facilitating adversarial learning between networks. This enables the generative model to create convincingly realistic fake samples from standard distribution data, and these deceptive samples, representing patterns of the original images, can be employed for subsequent training stages. On the other hand, GTNs leverage tractive learning between networks, enabling the generative model to analyze the changing patterns in the original time-series data. This analysis leads to the creation of new model-analyzed data, which, in turn, can generate high-precision artificial time series samples that closely resemble the variations in the original time series. All these novel data can be harnessed to enhance the performance of the subsequent predictive model.

Due to its reliance on principles from game theory, the GAN employs an objective function that embodies the adversarial process between the generative and discriminative models. This function is explicitly formulated as follows:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \tag{6}$$

In the GTN, the intention is to leverage the discriminative model to guide the training process of the generative model. Therefore, the objective function reflects the process in which the discriminative model guides the generative model. This objective function can be represented as follows:

$$\min_G \min_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [D(G(x)) - x]^2 + \mathbb{E}_{x \sim p_{data}(x)} [G(D(G(x))) - G(x)]^2 \tag{7}$$

From the objective function, it is evident that by optimizing the discriminative model, the generative model is further trained to generate new features are correlated with the original data distribution. The flowchart of the GTN can be found in Figure 7, and the algorithm pseudo-code is provided in Algorithm 1.

Algorithm 1 Minibatch Adam training of generative tractive network.

- Through time steps n sample onebatch of n features $[x_{t-n}, \dots, x_{t-1}]$ from time series.
 - for** number of training iterations **do**
 - Sample minibatch of m examples $\{x^{(1)}, \dots, x^{(m)}\}$ from data generating distribution $p_{data}(x)$.
 - Update netD by descending its gradient:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m [D(G(x^{(i)})) - x^{(i)}]^2$$
 - Use the updated netD update the generator by descending its gradient:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^m [G(D(G(x^{(i)}))) - G(x^{(i)})]^2$$
- end for**
-

The explanation of Figure 7 is as follows. First, construct autoregressive feature data that reflects the characteristics of multi-energy load sequences. The data are forwarded through the generative model netG to obtain the output of the generative model (OutputG). OutputG forwards through the tractive discriminative model netD, trained with the original time-series data as labels. OutputD, obtained by passing OutputG through the trained tractive discriminative model netD, serves as new data. This new data are used to train the generative model netG, considering OutputG as the label. Finally, observe whether the loss of the generative model (lossG) and the loss of the guidance identification model (lossD) approach zero. If the conditions are met, the training ends; otherwise, repeat the above training steps.

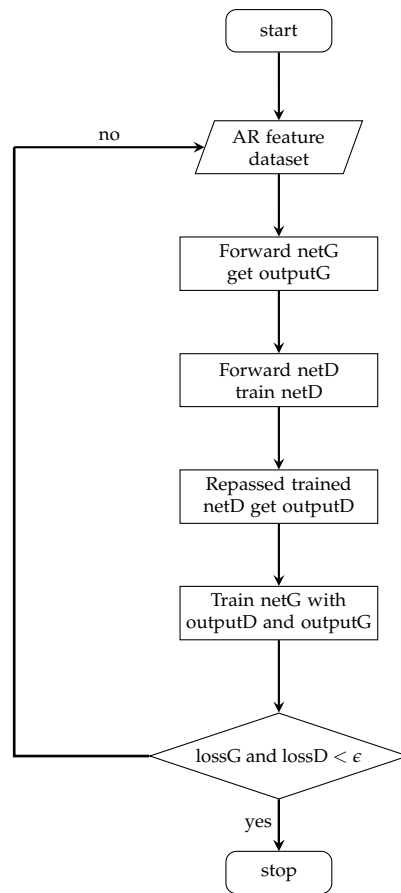


Figure 7. Flow chart of generative tractive network.

4.2. Feature Mining

For the generative model netG responsible for generating artificial features, this study employs the gated recurrent unit (GRU) sequence model [27]. This choice is attributed to the utilization of an additional input hidden state “h” of the sequence model during the GTN training process, which prevents the model from getting trapped in local minima during the initial phases of training. The discriminative tractive model netD is selected to be a multi-layer perceptron (MLP) with two hidden layers, each containing a variable number of neurons chosen from the range of [40, 30]. The loss function for the network is chosen as the mean squared error (MSE), often referred to as L2 loss.

After 250 epochs of iteration, the mean squared errors (MSEs) of netG and netD are presented in Table 1. The generated artificial features are shown in Figure 8.

Table 1. The loss of netG and netD.

Epoches	LossG	LossD
250	0.00020	0.00012

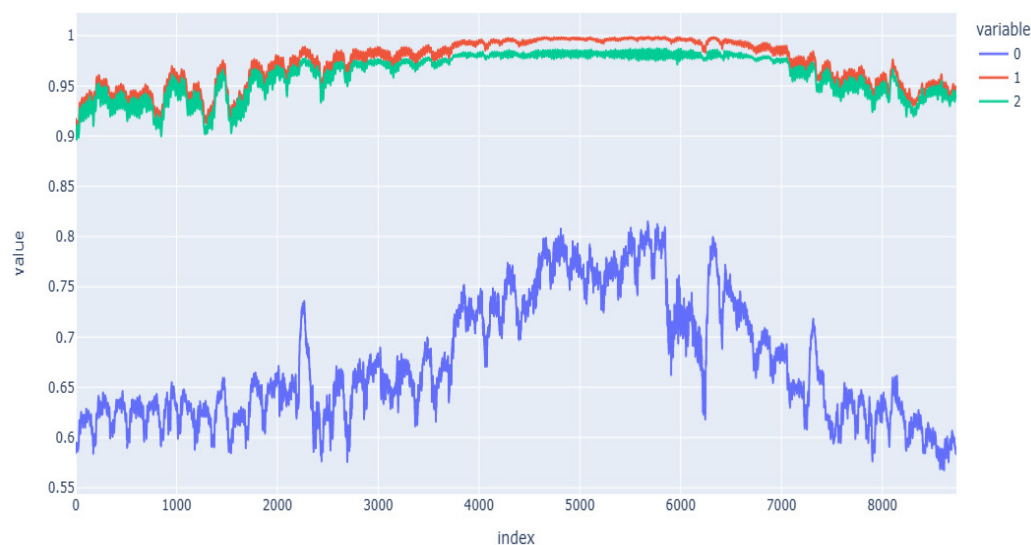


Figure 8. The artificial features generated by netG.

Regarding the artificial features obtained by netG from the data, these features are capable of generating the original data. In Figure 8, it can be observed that in features with significant fluctuation amplitudes, the network captures the overall variations of the original time series. Initially, there is a presence of seasonality, followed by an upward trend over time. After reaching a peak usage period, there is a subsequent downward trend. For the other two features with smaller fluctuation amplitudes, the model also captures a certain level of seasonality and exhibits a consistent downward trend, similar to the previous feature.

From Figure 9, it can be clearly seen that the artificially generated features capture the overall trend of the load. Additionally, it can be observed that both the artificially generated features and the average load curve appear very dense. This is because the load always follows a pattern of being high during the day and low at night, and there are no large-scale sudden changes within a period of time. This phenomenon indicates that the network has learned the seasonality of the load.

The generated artificial samples are depicted in Figure 10. From Figure 10, it can be observed that the artificial samples obtained through artificial features largely capture the seasonality and trends of the original time series. However, the model's performance is suboptimal during peak hours.

The newly added Table 2 shows the different statistical indicators between artificially generated samples and the original multi-energy time series. Regarding the averages, the generated electrical load and cooling load have averages of 0.305 and 0.375, respectively, which are close to the actual loads' averages of 0.312 and 0.389. However, the generated heating load has an average of 0.030, while the actual load has an average of 0.024, which is not ideal. This discrepancy is due to the data preprocessing method adopted in this study to reduce the impact of the heating load. Regarding the maximum values, the generated loads are 0.432, 0.789, and 0.053, while the actual loads are 0.581, 1.0, and 0.070, respectively. It can be observed that the network's performance is relatively poor in replicating peak values, indicating that the network struggles to learn extreme energy consumption patterns. Similar phenomena are observed in the representation of minimum values. The performance in maximum and minimum values also affects the variance, making the generated samples' variances smaller than those of the actual loads. Finally, the generated samples exhibit performance close to the actual loads across various quantiles, even for the heating load. Therefore, these samples can be utilized as new training samples for predictive models.

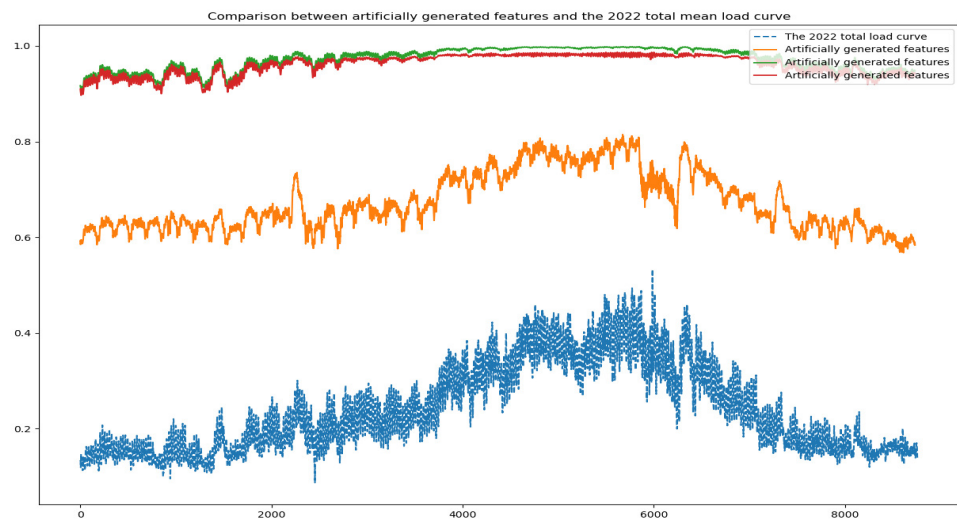


Figure 9. Comparison between artificially generated features and the 2022 total mean load curve.

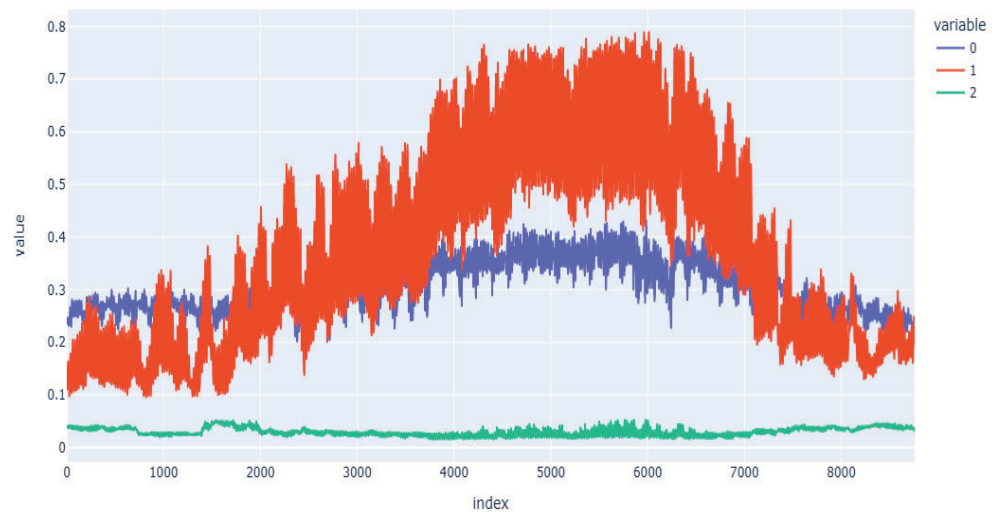


Figure 10. The artificial samples generated by netD.

Furthermore, because this paper uses the sequence model GRU as the generative model, the outputs of the generative model also include the hidden state outputs h . The GRU hidden state outputs of the generative model are shown in Figure 11. When combined with the unified unit 2022 multi-energy average curve in Figure 12, it can be observed that despite the differences in amplitude range, the two exhibit highly similar trends. This also indicates that the GRU generative model, in the process of analyzing the original data, captures the overall trends of the original multi-energy time series.

Table 2. Comparison between artificially generated samples and the original multi-energy time series.

	Mean	Std	Min	25%	50%	75%	Max
Artificial electricity	0.305	0.048	0.202	0.267	0.297	0.343	0.432
Original electricity	0.312	0.069	0.136	0.259	0.291	0.359	0.581
Artificial cool	0.375	0.191	0.096	0.205	0.336	0.521	0.789
Original cool	0.389	0.220	0.076	0.198	0.334	0.567	1.000
Artificial heat	0.030	0.008	0.016	0.024	0.028	0.036	0.053
Original heat	0.024	0.011	0.000	0.0162	0.019	0.031	0.070

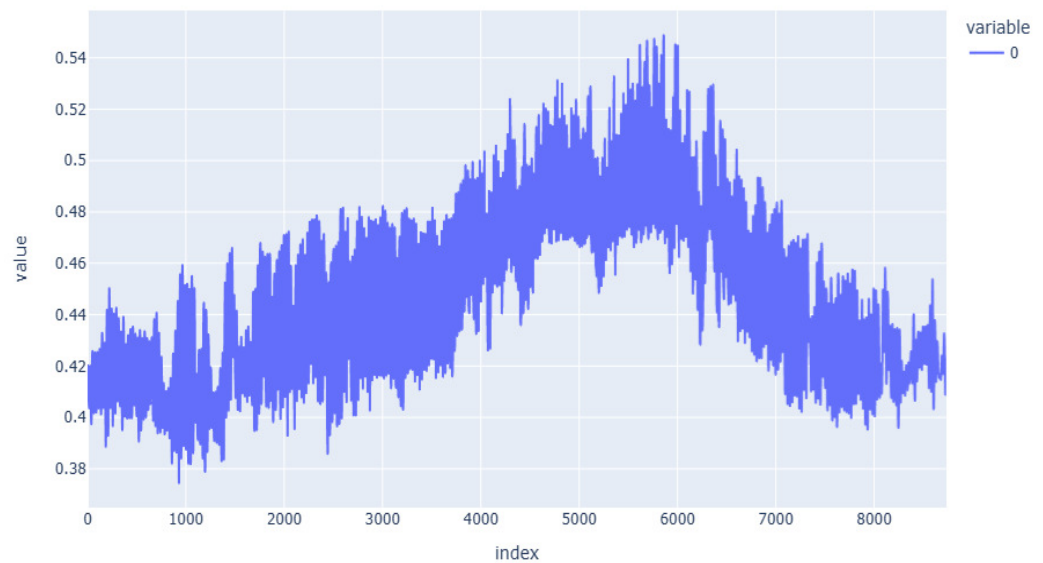


Figure 11. The output h of the generative model GRU hidden state.

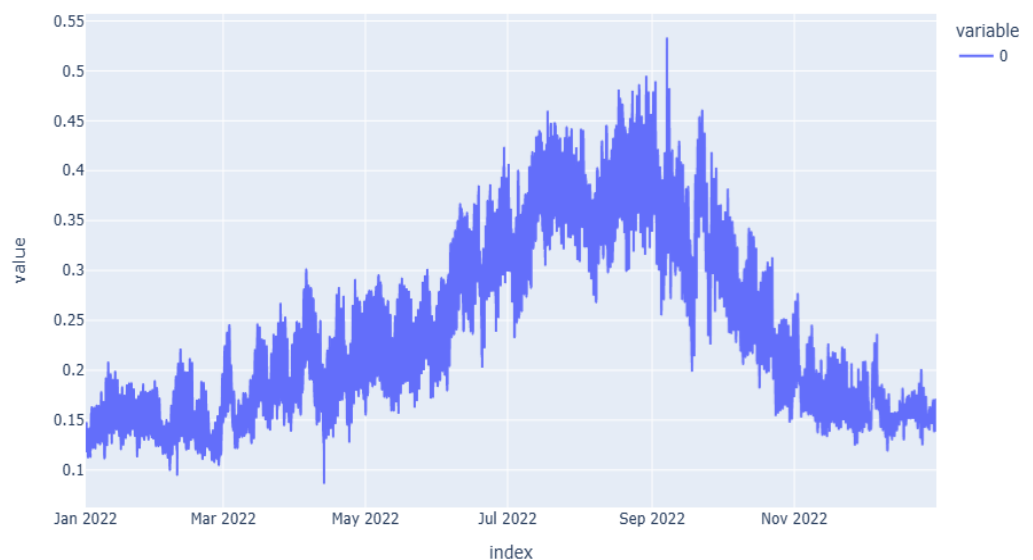


Figure 12. The unified unit multi-energy average curve in 2022.

5. Case Study

In the case study, this paper chooses to combine the GTN and GRU prediction model. Other compared prediction models include the GRU without the additional GTN component, non-sequential neural networks MLP, and other complex models proposed in previous literature. Additionally, this paper will analyze the impact of autoregressive data on the robustness of the prediction models. The specific experimental settings are as follows.

5.1. Experiment Setting

In order to validate the effectiveness of the proposed approach, this paper's prediction model will refrain from extensive optimization of most hyperparameters. All hyperparameters except for the autoregressive time step length will be randomly selected. The specific choices of parameters and model selection are outlined as follows:

- The autoregressive time step length is uniformly set to 24, as load variations exhibit at least a daily seasonality.

- The prediction model selects the sequence model gated recurrent unit (GRU) or a multi-layer perceptron (MLP) with two hidden layers.
- The dimension of the GRU sequence is set to 24.
- The MLP has two hidden layers with neuron counts of [40, 30], and ReLU activation functions are used. Given that load values are positive, the output layer’s activation function is chosen as the Sigmoid function.
- The batch size is fixed at 24.
- A single loss function, L1 loss (MAE), is selected. The overall loss function for multi-task target prediction is formulated as follows:

$$Loss = \sum_{i=1}^n w_i(t)L_i \tag{8}$$

Here, n represents the number of multi-tasks, and $w_i(t)$ and L_i denote the weight parameters and the loss function of the i -th task, respectively.

5.2. Multi-Objective Prediction Using GTN

The GTN has the capability to synthesize new features and samples, which provides more flexibility in data feature engineering. Depending on the composition of different time-series data, various artificially synthesized quantities can be employed to enhance the performance of prediction models. In this study, we conducted experiments using two different approaches for constructing time-series data: using only the historical data from the previous time step or utilizing the autoregressive feature data. The corresponding experiments are outlined below.

5.2.1. Predicting the Next Time Step Using Only the Historical Data from the Previous Time Step

Reference [15] proposed a comprehensive prediction method based on multi-task learning and bidirectional long short-term memory (MT-BiLSTM) networks for multi-source load data from the same system in the years 2016 to 2019. The performance of this method on the dataset used in this paper is shown in Table 3. Figure 13 demonstrates the fitting performance of this prediction method on the test set, showing significant deviations in the fitting for the cooling load. It can be observed that even though the model is relatively complex, it exhibits suboptimal performance on a small-scale dataset. This phenomenon can be attributed to the limited number and features of the samples.

Table 3. Experimental results.

	Electric Load		Cooling Load		Heating Load	
	Val MAPE	Test MAPE	Val MAPE	Test MAPE	Val MAPE	Test MAPE
GTN-GRU	2.53%	2.17%	2.77%	2.79%	4.16%	4.03 %
GRU	2.72%	2.32%	4.20%	3.08%	4.73%	4.62%
GTN-MLP	2.82%	2.32%	5.22%	4.46%	4.21%	4.41%
MLP	4.10%	4.48%	6.02%	5.25%	7.02%	8.21%
The predictive model in reference [15]	5.06%	5.45%	9.26%	6.58%	6.11%	5.01%

The highest accuracy in the same column is indicated in bold.

GRU utilizing the GTN and the results of comparative experiments are also presented in Table 3. The chosen error metric remains mean absolute percentage error (MAPE).

From the experimental results, it is evident that when only the historical sampling values of the previous time step are used as features, the GRU model shows an improvement in predictive generalization accuracy after undergoing data analysis with GTN. The improvement of the GTU-GRU model, in comparison to the GRU model, is particularly

pronounced for the cooling load prediction. This collectively indicates that the feature mining method based on GTN effectively enhances the predictive model performance. Additionally, with the introduction of new data, there is no noticeable fluctuation in the predictive accuracy of the multi-objective predictions. This further suggests that the GTN feature mining method contributes to an increased robustness in the predictive model. Furthermore, among all the model comparisons, the GTN-GRU model achieved the best prediction accuracy for all three energy sources, demonstrating the effectiveness of the approach presented in this paper. Figure 14 illustrates the fitting performance of the proposed GTN-GRU method on the test set, showing no significant deviations in the fitting for the three different energy loads.

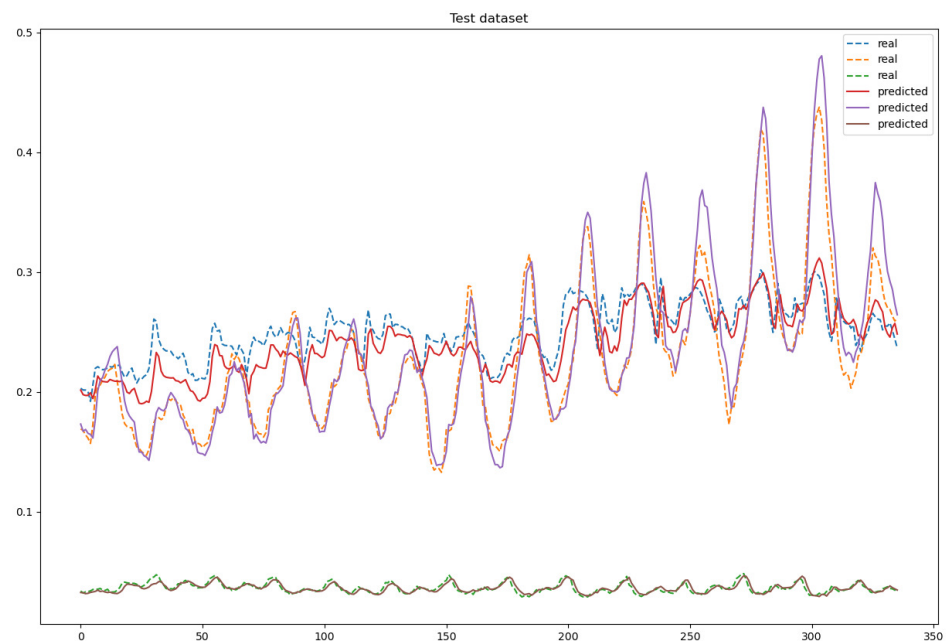


Figure 13. Performance of MT-BiLSTM in [15] on test set.

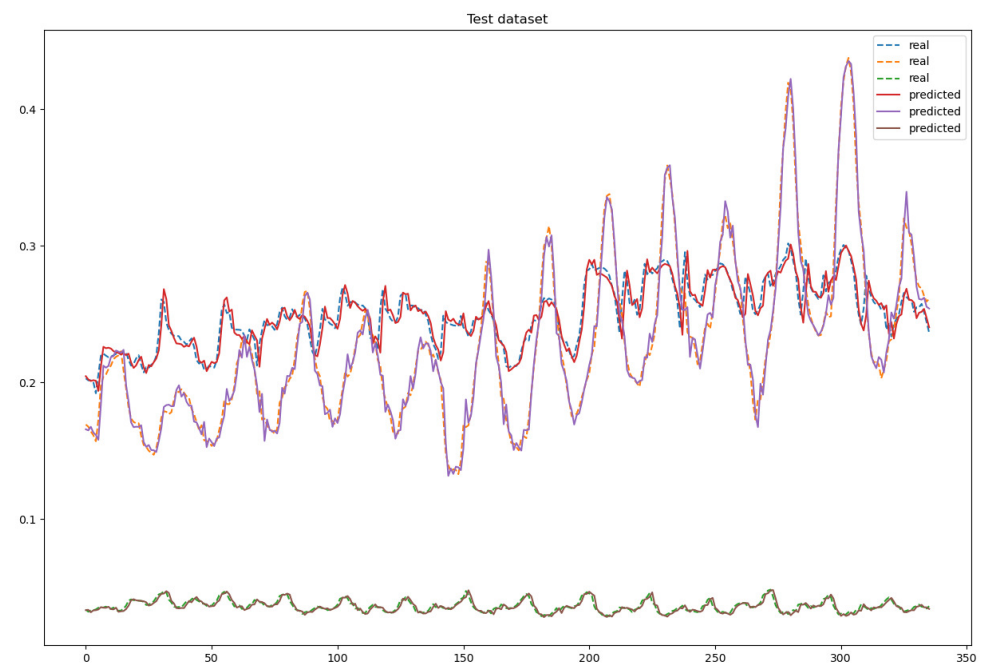


Figure 14. Performance of GTN-GRU on test set.

In addition, compared to the GRU model, the non-sequential MLP model performs poorly on the test set. However, by adding the GTN module, it can be observed that not only is the prediction accuracy of all three loads improved, but the prediction accuracy of the electric and thermal loads is also very close to the performance of the regular GRU model. This indicates that the proposed GTN model has an accuracy improvement effect even for non-sequential models.

5.2.2. Autoregressive Feature Data Prediction Model

Autoregressive feature data can be used not only for temporal feature mining by GTN but also for training prediction models. While examining the performance of the autoregressive feature data prediction model, this study also assesses the robustness of the model using the artificially generated features (add feature) and artificial samples (add sample) obtained through GTN. The specific experimental results are presented in Table 4.

From the experimental results, it can be observed that using the original dataset alone, the autoregressive feature construction model exhibits decent accuracy performance, aligning with the conclusion that autoregressive features can better capture historical data patterns. However, as shown in Table 4, when different amounts of artificial data are added, hardly any model exhibits stable performance in terms of generalization metrics. The performance of the autoregressive feature construction model starts to become unstable. This is attributed to the significant use of historical data as features, which violates the principle of i.i.d data. Consequently, the model's robustness is compromised, and fluctuations in prediction results arise when the data changes. In this paper, autoregressive data are applied in the generative model to produce analytical results of the original time series. These results are utilized as new data features for training predictive models. Furthermore, this paper only predicts the next moment using the current time point. The theoretical basis for improving model robustness through artificially generated data can be found in reference [23]. Overall, this study opted for trading computational cost for improved prediction accuracy and ensured model robustness. Given that the training dataset in this study is relatively small, the increase in computational cost is acceptable.

Table 4. AR feature data experimental results.

	Electric Load		Cooling Load		Heating Load	
	Val MAPE	Test MAPE	Val MAPE	Test MAPE	Val MAPE	Test MAPE
AR feature MLP	2.66%	2.26%	3.20%	3.15%	5.5%	5.17%
AR feature MLP (add sample)	2.73%	2.38%	2.99%	3.20%	4.05%	3.75%
AR feature MLP (add feature)	2.89%	3.24%	2.68%	3.16%	9.89%	8.25%
AR feature GRU	2.69%	2.45%	3.94%	2.91%	6.02%	3.51%
AR feature GRU (add sample)	2.93%	2.22%	4.78%	3.36%	12.58%	3.70%
AR feature GRU (add feature)	5.09%	3.30%	4.91%	4.12%	9.52%	10.31%

The highest accuracy in the same column is indicated in bold.

6. Conclusions

This paper proposes a feature mining approach based on a generative tractive network (GTN) to address the multi-objective time-series prediction problem in integrated energy systems. By constructing autoregressive feature data to express time-varying patterns, such data, although directly applicable to predictive models, can lead to decreased robustness due to severe violations of the i.i.d condition. Therefore, they are not suitable for combining with other feature engineering methods. However, data with such evident sequential characteristics can be used for training generative models. Based on multi-task learning, the GTN model proposed in this paper analyzes multi-objective time series in integrated energy systems. The analysis results generated by the generative model are used by the tractive discriminative model to determine whether they can map to the original time series.

Ultimately, the generative model and the tractive discriminative model approximate an inverse mapping relationship, and the artificially synthesized data generated by both are highly related to the original time series. After adding artificially synthesized data, the performance and robustness of the predictive model are both improved.

Author Contributions: Methodology, Z.Z.; software, Z.Z.; validation, Z.Z.; formal analysis, Z.Z.; investigation, Z.Z.; resources, Z.Z.; data curation, Z.Z.; writing—original draft, Z.Z.; visualization, Z.Z.; supervision, Z.W.; project administration, Z.W.; funding acquisition, Z.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China (61973070, 62373089), the Nature Science Foundation of Liaoning Province, China under Grant 2022JH25/10100008.

Data Availability Statement: This data are available in <http://cm.asu.edu/>, accessed on 22 May 2023.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Zhu, J.; Dong, H.; Zheng, W.; Li, S.; Huang, Y.; Xi, L. Review and prospect of data-driven techniques for load forecasting in integrated energy systems. *Appl. Energy* **2022**, *321*, 119269. [[CrossRef](#)]
- Lee, C.M.; Ko, C.N. Short-term load forecasting using lifting scheme and ARIMA models. *Expert Syst. Appl.* **2011**, *38*, 5902–5911. [[CrossRef](#)]
- Ding, S.; Zhang, H.; Tao, Z.; Li, R. Integrating data decomposition and machine learning methods: An empirical proposition and analysis for renewable energy generation forecasting. *Expert Syst. Appl.* **2022**, *204*, 117635. [[CrossRef](#)]
- Zhang, W.; Robinson, C.; Guhathakurta, S.; Garikapati, V.M.; Dilkina, B.; Brown, M.A.; Pendyala, R.M. Estimating residential energy consumption in metropolitan areas: A microsimulation approach. *Energy* **2018**, *155*, 162–173. [[CrossRef](#)]
- Robinson, C.; Dilkina, B.; Hubbs, J.; Zhang, W.; Guhathakurta, S.; Brown, M.A.; Pendyala, R.M. Machine learning approaches for estimating commercial building energy consumption. *Appl. Energy* **2017**, *208*, 889–904. [[CrossRef](#)]
- Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2023**, arXiv:1706.03762.
- Meng, A.; Wang, P.; Zhai, G.; Zeng, C.; Chen, S.; Yang, X.; Yin, H. Electricity price forecasting with high penetration of renewable energy using attention-based LSTM network trained by crisscross optimization. *Energy* **2022**, *254*, 124212. [[CrossRef](#)]
- Arastehfar, S.; Matinkia, M.; Jabbarpour, M.R. Short-term residential load forecasting using Graph Convolutional Recurrent Neural Networks. *Eng. Appl. Artif. Intell.* **2022**, *116*, 105358. [[CrossRef](#)]
- Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *arXiv* **2014**, arXiv:1406.2661.
- Lei, Y.; Wang, D.; Jia, H.; Chen, J.; Li, J.; Song, Y.; Li, J. Multi-objective stochastic expansion planning based on multi-dimensional correlation scenario generation method for regional integrated energy system integrated renewable energy. *Appl. Energy* **2020**, *276*, 115395. [[CrossRef](#)]
- You, M.; Wang, Q.; Sun, H.; Castro, I.; Jiang, J. Digital twins based day-ahead integrated energy system scheduling under load and renewable energy uncertainties. *Appl. Energy* **2022**, *305*, 117899. [[CrossRef](#)]
- Zhou, B.; Meng, Y.; Huang, W.; Wang, H.; Deng, L.; Huang, S.; Wei, J. Multi-energy net load forecasting for integrated local energy systems with heterogeneous prosumers. *Int. J. Electr. Power Energy Syst.* **2021**, *126*, 106542. [[CrossRef](#)]
- Wang, X.; Wang, S.; Zhao, Q.; Wang, S.; Fu, L. A multi-energy load prediction model based on deep multi-task learning and ensemble approach for regional integrated energy systems. *Int. J. Electr. Power Energy Syst.* **2021**, *126*, 106583. [[CrossRef](#)]
- Zhao, H.; Guo, S. Uncertain Interval Forecasting for Combined Electricity-Heat-Cooling-Gas Loads in the Integrated Energy System Based on Multi-Task Learning and Multi-Kernel Extreme Learning Machine. *Mathematics* **2021**, *9*, 1645. [[CrossRef](#)]
- Guo, Y.; Li, Y.; Qiao, X.; Zhang, Z.; Zhou, W.; Mei, Y.; Lin, J.; Zhou, Y.; Nakanishi, Y. BiLSTM Multitask Learning-Based Combined Load Forecasting Considering the Loads Coupling Relationship for Multienergy System. *IEEE Trans. Smart Grid* **2022**, *13*, 3481–3492. [[CrossRef](#)]
- Li, C.; Li, G.; Wang, K.; Han, B. A multi-energy load forecasting method based on parallel architecture CNN-GRU and transfer learning for data deficient integrated energy systems. *Energy* **2022**, *259*, 124967. [[CrossRef](#)]
- Chung, W.H.; Gu, Y.H.; Yoo, S.J. District heater load forecasting based on machine learning and parallel CNN-LSTM attention. *Energy* **2022**, *246*, 123350. [[CrossRef](#)]
- Tan, Z.; De, G.; Li, M.; Lin, H.; Yang, S.; Huang, L.; Tan, Q. Combined electricity-heat-cooling-gas load forecasting model for integrated energy system based on multi-task learning and least square support vector machine. *J. Clean. Prod.* **2020**, *248*, 119252. [[CrossRef](#)]
- Niu, D.; Yu, M.; Sun, L.; Gao, T.; Wang, K. Short-term multi-energy load forecasting for integrated energy systems based on CNN-BiGRU optimized by attention mechanism. *Appl. Energy* **2022**, *313*, 118801. [[CrossRef](#)]

20. Wang, S.; Wang, S.; Chen, H.; Gu, Q. Multi-energy load forecasting for regional integrated energy systems considering temporal dynamic and coupling characteristics. *Energy* **2020**, *195*, 116964. [[CrossRef](#)]
21. Zhao, J.; Liu, X. A hybrid method of dynamic cooling and heating load forecasting for office buildings based on artificial intelligence and regression analysis. *Energy Build.* **2018**, *174*, 293–308. [[CrossRef](#)]
22. Pearl, J.; Mackenzie, D. *The Book of Why: The New Science of Cause and Effect*; Penguin Books: Harlow, UK, 2019.
23. Xing, Y.; Song, Q.; Cheng, G. Why Do Artificially Generated Data Help Adversarial Robustness. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 954–966.
24. University AS. Campus Metabolism. Available online: <http://cm.asu.edu/> (accessed on 22 May 2023).
25. Raschka, S.; Liu, Y.; Mirjalili, V.; Dzhulgakov, D. *Machine Learning with PyTorch and Scikit-Learn: Develop Machine Learning and Deep Learning Models with Python*; Expert Insight; Packt Publishing: Birmingham, UK, 2022.
26. Box, G.; Jenkins, G. *Time Series Analysis: Forecasting and Control*; Holden-Day Series in Time Series Analysis and Digital Processing; Holden-Day: San Francisco, CA, USA, 1976.
27. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *arXiv* **2014**, arXiv:1412.3555.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.