applied sciences

MDPI

Article

Channel Modeling Based on Transformer Symbolic Regression for Inter-Satellite Terahertz Communication

Yuanzhi He *, Biao Sheng and Zhiqiang Li

Institute of Systems Engineering, Academy of Military Sciences, Beijing 100141, China; shengb11@126.com (B.S.); lzq_000@126.com (Z.L.)
* Correspondence: he_yuanzhi@126.com

Abstract: Channel modeling is crucial for inter-satellite terahertz communication system design. The conventional method involves manually constructing a mathematical channel model, which is labor-intensive, and using a neural network directly as a channel model lacks interpretability. This paper introduces a channel modeling approach based on symbolic regression. It is the first time that using transformer neural networks as the implementation tool of symbolic regression to generate the mathematical channel model from the channel data directly. It can save manpower and avoid the interpretability issue of using neural networks as a channel model. The feasibility of the proposed method is verified by generating a free space path loss model from simulation data in the terahertz frequency band.

Keywords: symbolic regression; channel modeling; terahertz; artificial intelligence

check for updates

Citation: He, Y.; Sheng, B.; Li, Z. Channel Modeling Based on Transformer Symbolic Regression for Inter-Satellite Terahertz Communication. *Appl. Sci.* **2024**, *14*, 2929. https://doi.org/10.3390/app14072929

Academic Editor: Gordana Jovanovic Dolecek

Received: 17 January 2024
Revised: 15 March 2024
Accepted: 23 March 2024
Published: 30 March 2024

Copyright: © 2024 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

The vigorous development of satellite communication meets the needs of social development. In recent years, with the rise of low earth orbit constellations, the demand for data exchange between satellites is increasing, thus there is a great demand for spectrum resources. The L, S, C, X, Ku, and Ka frequency bands for traditional satellite communication are very scarce. As an undeveloped band of the electromagnetic spectrum, the terahertz band has a frequency between 0.1~10 THz, which has a high spectrum bandwidth to realize high-speed data transmission, and has great application value in satellite communication. At present, International Telecommunication Union has completed the frequency division of satellite services in the frequency range of 100~275 GHz and has made a simple division of the terahertz frequency band above 275 GHz. At the same time, the development of terahertz devices is progressing steadily. The terahertz frequency band has been preliminarily qualified for application in satellite communication.

The modeling of inter-satellite terahertz channel is the basis for realizing the application of inter-satellite terahertz wireless communication. The channel through which terahertz waves are transmitted from the transmitter antenna to the receiver antenna is the terahertz wireless channel, and the characteristics of the channel determine the performance of satellite communication systems. The terahertz channel model is the foundation for the design and optimization of terahertz communication systems. Channel modeling is based on data measured by existing channel measurement platforms, using mathematical formulas to characterize various parameters of the channel.

There have been many researches on the modeling of terahertz channel in terrestrial communication [1–4]. Tian et al. [1] provided a detailed explanation of the characteristics of ground terahertz channels. The terahertz frequency is large and the wavelength is short, so the terahertz channel has different channel characteristics compared with the low-frequency channel. As terahertz travels through the atmosphere, molecules such as water vapor, clouds, ice crystals, and dust increase path losses. Han et al. [2] analyzed the existing terahertz channel models and introduced channel simulators based on these models. Including

NYUSIM, Cloud RT, EDX Advanced Promotion, etc. These models are mainly applied in ground communication scenarios such as indoor offices and outdoor environments.

However, to our knowledge, there is currently a lack of research on channel modeling for inter-satellite terahertz communications, therefore it is necessary to conduct relevant research.

In addition, in terms of channel modeling methods, channel modeling methods include deterministic modeling and statistical modeling. The deterministic channel modeling method is based on the analysis of optical and electromagnetic propagation theories in current application scenarios to establish a wireless channel model. Priebe et al. [5] used ray tracing technology to model a 300 GHz indoor environment and used a free space path loss formula to model large-scale fading of line-of-sight links. For small-scale fading, the line-of-sight links delay is modeled as $\tau = \frac{d}{c}$, recursively calculating the delay of each order of reflection path. The line-of-sight links phase is modeled as the first-order function of the delay $\varphi = -2\pi f\tau$, the phase on the reflection path is uniformly modeled within -180° and 180°. And model the horizontal angle of arrival (AOA) uniformly, and calculate the horizontal angle of departure (AOD) by adding a difference value on the AOA, which is equal to a multiple of 180°. The advantage of the deterministic channel modeling method is that it does not require actual measurement, but its disadvantage is that it requires very detailed application scenario information and high computational complexity. Sometimes, in order to make the modeled channel model more concise and practical, deterministic channel modeling methods may have certain trade-offs and idealization of the parameters that affect the channel during modeling. Therefore, some scholars use statistical modeling methods for channel modeling.

Statistical channel modeling uses a measurement platform to measure channel data in actual application scenarios and fits the actual data to obtain the empirical distribution and statistical characteristics of each channel parameter. Finally, the channel is reconstructed based on statistical characteristics.

He et al. [6] conducted channel measurements from 220 GHz to 340 GHz using a vector network analyzer and proposed a propagation channel model for the terahertz band. It is represented by a logarithmic model:

$$\alpha_{\text{trans}} = 20\lg\left(\frac{c}{4\pi f}\right) - A_{PET} - 20\lg(d) \tag{1}$$

According to the measurement results, it was found $A_{PET} = 1.97\text{dB}$.

Traditional statistical methods rely on manually constructing mathematical channel model from data, which is time-consuming. In order to solve this problem, some researchers use neural networks to directly fit the channel model. On the one hand, for the existing wireless channel models, it is only necessary to train the neural network with the data generated by the corresponding models, so that the neural network can approximate the actual channel model under the minimum mean square error criterion. On the other hand, for wireless channels without channel models, neural network-based channel modeling uses measured data for training and does not need to determine the propagation path of electromagnetic waves, so it is not constrained by the environment and is more suitable for various complex scenarios. When the actual wireless channels are nonlinear/non-stationary, neural networks have a good performance in simulating nonlinear systems. Neural networks can train high-dimensional nonlinear data, and solve many problems that are difficult to be solved by traditional modeling methods.

Bai et al. [7] proposed a 3D MIMO indoor channel modeling method targeting the millimeter wave frequency band. Based on the convolutional neural network model, the input data are the coordinates of the transmitter and receiver, and the output characteristic parameters include the received power, delay, transmission azimuth, transmission elevation, arrival azimuth, and arrival elevation.

Ferreira et al. [8] used neural networks to improve the prediction of outdoor signal strength in the ultra-high frequency (UHF) band. The diffraction loss and transmitted signal strength are fed into the neural network and the strength of the received signal will

be output at the output layer. The results show that the neural networks can improve the prediction of outdoor signal strength in the UHF band.

However, these methods of using neural networks as a channel model make it difficult to reveal the mathematical characteristics and physical mechanism of the channel because of the interpretability issue of neural networks.

Xue et al. [9] noted the interpretability issue of using neural networks as a channel model and proposed a scheme to use causal neural networks for channel modeling. However, this method still uses neural networks directly as the channel model, although it enhances the interpretability of neural networks, it is not intuitive enough. In order to solve the problem of insufficient interpretability of neural network channels, Lee et al. [10] first used channel data to train the neural networks as a channel model, and then used a genetic algorithm to generate a symbolic regression formula from the neural network channel model. This is the indirect use of the symbolic regression method to generate the channel model. This indirection may be unnecessary.

Given the analysis of the above factors, this paper proposes a transformer symbolic regression-based inter-satellite terahertz channel modeling method which is based on the symbolic regression method PhySO [11]. It uses a transformer neural network as a tool to directly fit the mathematical channel model from the measured channel data, avoiding the laborious problem of establishing a mathematical channel model by traditional statistical methods and the interpretability issue caused by using a neural network as a channel model, and can reveal the mathematical relationship and physical mechanism between channel parameters. Figures 1 and 2 show the difference between using neural networks directly as channel model and using deep neural network as a symbolic regression tool to generate channel model.

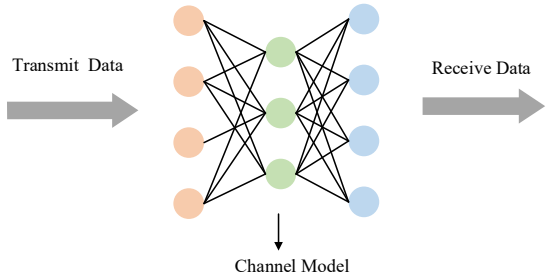


Figure 1. Deep neural network as channel model.

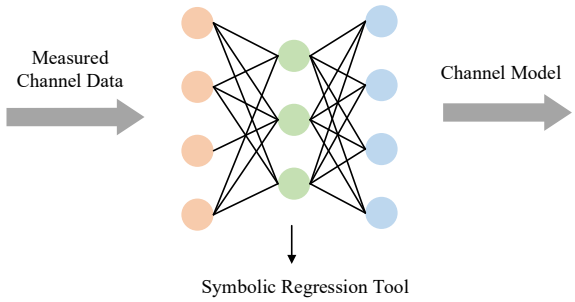


Figure 2. Deep neural network as a symbolic regression tool.

The main contribution of this paper are:

- To the best of our knowledge, it is the first time to establish a mathematical channel model directly from data by using a symbolic regression method. As a new channel modeling method, it may help researchers to establish a channel model easily.
- The PhySO method is improved to be more suitable for channel modeling tasks.

The method proposed in this article may have important application prospects in the field of channel modeling. Although this article takes the terahertz frequency band as an

example, it has a wide range of applications and can be used as a data-fitting tool for other frequency bands and more complex communication scenarios.

This paper is organized as follows: Section 1, introduces the important role of channel modeling in the future inter-satellite terahertz communication systems, expounds on the shortcomings of the traditional channel modeling scheme and the use of neural network as the channel model scheme, and proposes a channel modeling scheme based on symbolic regression. Section 2, introduces the symbolic regression method, especially the PhySO symbolic regression method in detail, and the corresponding improvement to PhySO, which makes it more suitable for the channel modeling task. Section 3, simulates the fitting effect of the improved PhySO symbolic regression method on the free space path loss model and verifies the feasibility of the proposed method. In Section 4, it summarizes the characteristics of the proposed method and points out the next research plan.

2. Transformer Symbolic Regression

2.1. Symbolic Regression

Symbolic regression is a technique aimed to automatically discover mathematical expressions or functions from data sets [12].

Set target dataset $X = \{x_1, x_2, x_3, \ldots, x_n\}$ and label $Y = \{y_1, y_2, y_3, \ldots, y_n\}$, symbolic regression aims to find a function $f(\cdot)$ so that $f = \text{argmin}(eval(f(X)))$ where $eval(\cdot)$ indicates the evaluation of formula $f(\cdot)$. For example, $eval(\cdot) = [f(X) - Y]^2$.

Symbolic regression is considered an extension of traditional regression methods, traditional regression methods need to assume that the data follows a polynomial distribution or some other distribution while symbolic regression doesn't have to. Symbolic regression can automatically discover nonlinear and higher-order relationships in the data, and can also be applied to multivariable problems to discover interactions between multiple variables. It can be used to uncover underlying patterns behind data to help researchers find real models and mechanisms.

Symbolic regression is mainly realized by genetic algorithm and deep reinforcement learning algorithm.

Traditional symbolic regression algorithms mainly use genetic algorithms to search for feasible mathematical expressions.

The specific steps of using the genetic algorithms to search for the optimal formula are as follows:

1. Set the initial space of arithmetic variables, operations and end conditions during the running process;
2. Initial formula set as population;
3. Evaluate the formulas based on evaluation $eval(\cdot)$ such as the mean square error between the predicted values of formulas and the data labels;
4. Generate a new population by replication, crossover, and mutation operations;
5. Repeat steps 3 and 4 until the end condition is met, and sort the generated formula set based on the evaluations to choose the best formula.

Symbolic regression based on genetic algorithm has some problems, it cannot take advantage of the inherent characteristics of the dataset to search for suitable formula, so the search process will be too long, the search is too inefficient, and it is easy to fall into local optimal.

In recent years, deep reinforcement learning has made great progress in the field of optimization solutions. Many studies have applied deep reinforcement learning to solve symbolic regression problems. This kind of method can search symbols from the symbol space (the space of variables and operations), construct a mathematical expression from the symbols, and get a reward according to the fitness of the mathematical expression to optimize the strategy function. Deep reinforcement learning can achieve very good results in symbolic regression problems [13].

Deep learning-based symbolic regression methods often model formulas as binary trees, the nodes of the tree are called symbols which are divided into variables in green

and operations in blue of mathematical, as shown in Figure 3, the binary tree can be converted into mathematical formulas by depth-first search of mid-order traversal, and it is $\cos(x \times y) + \tan(x) / \ln(x + y)$. Therefore, the problem of symbolic space exploration is regarded as the generation problem of a formula binary tree.

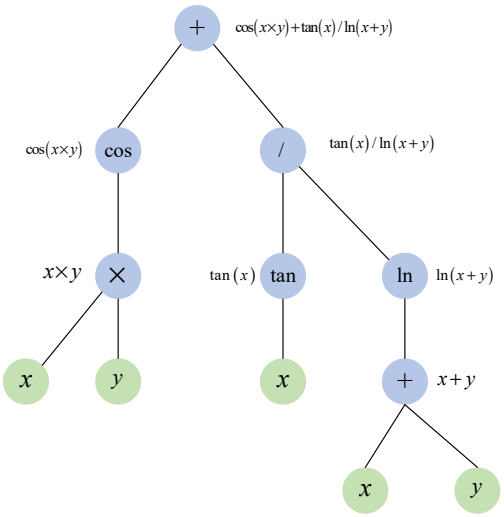


Figure 3. Binary tree of formula.

2.2. PhySO

Thanks to Tenachi et al., they proposed PhySO [11], a powerful deep reinforcement learning-based symbolic regression method. It builds a mathematical expression starting from the most basic physical units and automatically detects and corrects combinations of symbols that may lead to violations of physical unit constraints, ensuring that unit correctness is maintained throughout all computations. This avoids generating irrational mathematical expressions and greatly reduces the size of the expression search space, so the search process of PhySO is more efficient and can find the best solution faster.

In the implementation, PhySO models mathematical expressions as binary trees, where variables and coefficients are terminal nodes, operation symbols are non-terminal nodes, monadic operators have only 1 child node, and binocular operators have 2 child nodes. Variables and operators are called tokens in PhySO, and the space they make up is called Library, for example, Library = {a, b, c, +, −, /}. Tokens in the Library such as "a, b, c, +, −, /" are encoded by one-hot for subsequent processing by the neural network.

As a deep reinforcement learning method, PhySO sets the observations as parent nodes and their units, sibling nodes and their units, previous node and its unit, dangling nodes and the unit of the current node, and the initial observations are all zero tensor. The action is to select the tokens in the Library to build the mathematical expression. By prioritizing the output of the action by the neural network and performing the mask operation, it masks out unnecessary actions to generate better mathematical expressions. The reward is:

$$R = \frac{1}{1 + NRMSE} \tag{2}$$

$$\sigma_y = \sqrt{\frac{1}{N - \delta_N} \sum_{i=1}^{N} \left(y_i - \frac{1}{N} \sum_{i=1}^{N} y_i \right)^2} \tag{3}$$

$$NRMSE = \frac{1}{\sigma_y} \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - f(x_i))^2} \tag{4}$$

where R is the reward, σ_y is the standard deviation of the target value, N is the number of sampling points, δ_N is Bessel's correction and $\delta_N = 1$, x_i, y_i are data points and target values, and $f(\cdot)$ is the function generated by the neural network.

The optimization strategy used by PhySO is risk-seeking policy gradient [14], entropy regularization [15], and Adam optimizer is adopted. When the reward is big enough and the expression is meaningful, the algorithm will stop iterating and output the corresponding result.

The network architecture used by PhySO is Long Short-Term Memory (LSTM) which is a classical neural network architecture for processing sequence data. After the observations data is input into the LSTM network, the network will output the probability distribution of each action, adjust the probability distribution of the action by masking the actions that do not meet the constraint conditions, and then sample the action according to the probability. Then, based on the action, the new mathematical expression is obtained, and the observations are updated. It will be repeated until the final mathematical expression is obtained.

By deep reinforcement learning technology, the PhySO method can adaptively adjust learning strategies, and automatically select the most appropriate mathematical expression to describe data by using physical constraints and avoid meaningless symbol combinations. This method can greatly improve the search efficiency and reliability of the model, and can better adapt to different types of physical data.

2.3. Improved PhySO Algorithm

In the PhySO project, the authors used LSTM as a neural network architecture, and only the optimal 5% of candidate solutions were rewarded. However, LSTM has the problem of poor parallel performance, and each LSTM cell has four fully connected layers, if the LSTM network is very deep, the computational load will be large and time-consuming. And only the optimal 5% candidate solution is rewarded, when the channel mathematical model is relatively complex, such as when there are mixed operations of logarithm, exponent, fraction, and trigonometric function in mathematical expressions, there are problems of poor exploration performance and slow convergence. In addition, in the process of generating a channel model with PhySO, multilayer fractions, and multilayer exponents often appear, which are not common in the channel model. In order to solve these problems, the PhySO method is improved to make it more suitable for channel modeling tasks.

In this paper, the LSTM architecture is changed to transformer architecture [16] to increase the parallel performance and feature extraction capability of the algorithm. The self-attention mechanism in the transformer architecture enables the calculation of each time step to only rely on the input vector, thus achieving completely parallel computation. Moreover, the self-attention mechanism can directly calculate the dependency relationship between any two positions in the sequence, making the model better able to capture long-distance dependencies. The structure of the transformer model is very flexible and can be adjusted according to the needs of specific tasks, such as increasing or decreasing the number of layers and adjusting the number of attention mechanism heads. Its architecture diagram, as shown in Figure 4, includes the input layer, transformer layer, and output layer. Both the input and output layers are linear layers, and the transformer architecture used contains only the Encoder part, which is used to extract the features of mathematical expressions.

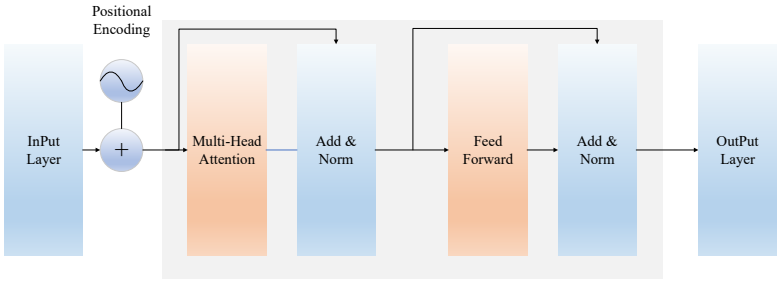


Figure 4. Network architecture diagram.

In the process of exploration, the optimal 5% candidate solution is changed to the optimal candidate solution decreasing from 10–5% to increase the exploration effect of the model.

According to the channel model expression, the symbol space is redefined as $\left\{+, -, \times, /, \exp, \log, \frac{1}{x}, x^2, \sqrt{x}\right\}$. By reducing unnecessary operators, search complexity can be reduced and convergence speed can be accelerated.

Modify the config file of the PhySO project to reduce the occurrence probability of multiple fractions and exponents. Because in the common channel model, it is unusual to see the nested fraction and exponential function. By reducing the nesting of fractional, exponential, and logarithmic operations, the complexity of algorithm search can also be reduced, resulting in a faster search for the optimal expression.

The logarithmic function used in the PhySO project is changed to the logarithmic form with base 10, which is more consistent with the dB definition in the channel model.

3. Experiment

3.1. Terahertz Channel Model

Moldovan et al., based on scattering theory and ray-tracing technique, proposed a deterministic large-scale fading model suitable for terahertz frequency bands. It deterministically models the line-of-sight loss as the sum of free space path loss and molecular absorption loss [17]:

$$A(f,d) = A_{spread}(f,d) + A_{abs}(f,d) \tag{5}$$

where $A(f,d)$ represents terahertz deterministic large-scale fading, $A_{spread}(f,d)$ represents free space path loss, and $A_{abs}(f,d)$ represents molecular absorption loss.

Considering that the inter-satellite terahertz communication scenario is located in space, the molecular absorption loss is very small, so only the free space path loss is considered for the inter-satellite terahertz channel modeling.

Free space path loss is a classical model of wireless channel transmission, which plays an important role in satellite communication simulation. In the case of an ideal omnidirectional antenna, the free space path loss can be expressed as the ratio of the transmitted and received power of the antenna.

$$\frac{P_t}{P_r} = \frac{(4\pi f d)^2}{c^2} \tag{6}$$

The logarithmic form is:

$$\begin{aligned} L_{dB} &= 10\lg\left(\frac{P_t}{P_r}\right) \\ &= 10\lg\left(\frac{(4\pi f d)^2}{c^2}\right) \\ &= 20[\lg\left(\frac{4\pi}{c}\right) + \lg(f) + \lg(d)] \\ &\approx 92.44 + 20\lg(f) + 20\lg(d) \end{aligned} \tag{7}$$

where, P_t is the transmitting power, in the unit w, P_r is the receiving power, in the unit w, f is the signal frequency, in the unit GHz, d is the distance, in the unit km. when calculating the final value of L_{dB}, units shall not be included.

3.2. Experimental Process

As shown in Figure 5, based on the PhySO project, the experimental process of this article is mainly divided into the following steps:

1. Input the measured channel data. In this experiment, the data was sampled from the free space path loss Formula (7);
2. Initialize the environment, including batch size, library, priority of operations, dataset, etc.;
3. Initialize the deep neural network, including its size, learning rate, etc.;

4. Initialize the observations which are composed of parent nodes and their units, sibling nodes and their units, previous node and its unit, dangling nodes and the unit of the current node;
5. Input the observations into the neural network model to calculate the action, and use the action to form the channel model expression;
6. According to the reward Formula (2), calculate the reward using the sampled dataset;
7. After receiving the reward, use the risk-seeking policy gradient [14], and entropy regulation [15] method to calculate the loss;
8. Backpropagation loss, using Adam optimizer to optimize neural network and update the observations;
9. Repeat 5–8 until the reward value reaches the set value or the maximum number of iterations is reached.

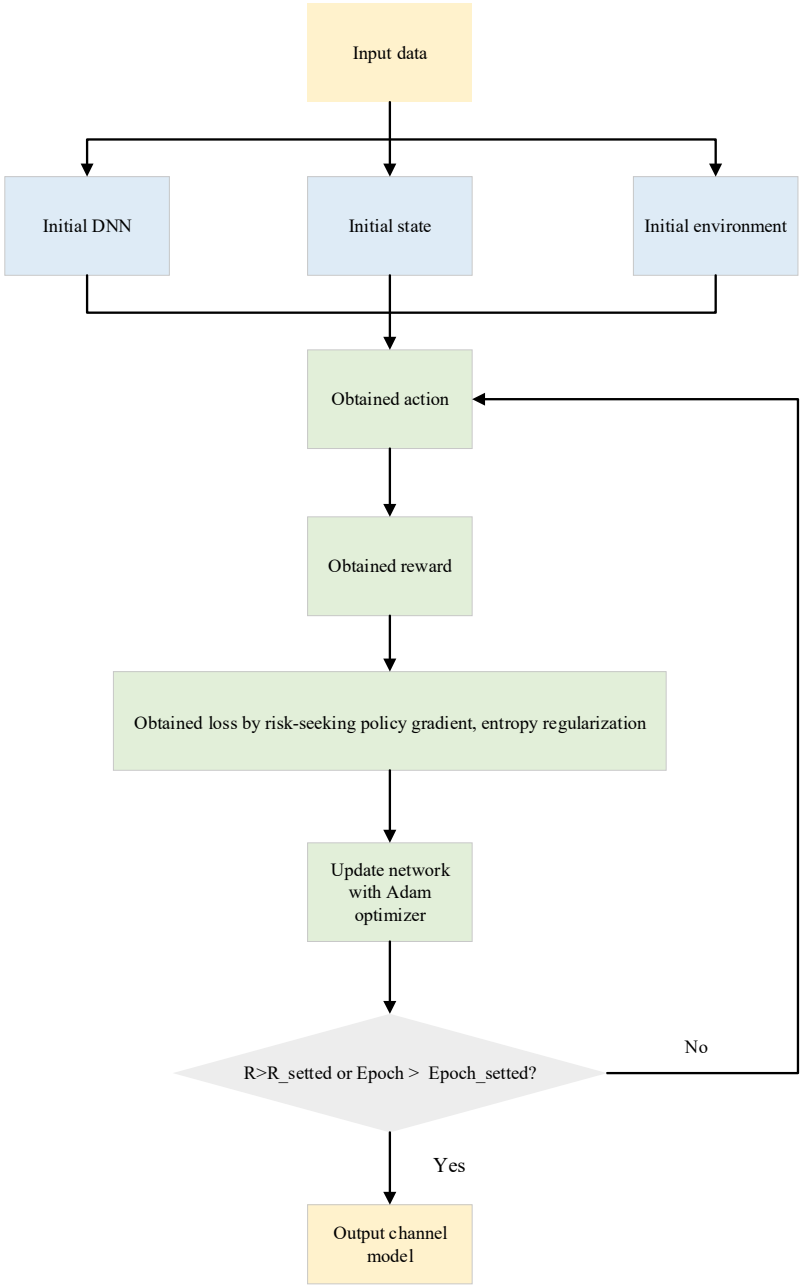


Figure 5. Experimental Process.

3.3. Experimental Parameter

In this paper, the terahertz frequency band is selected for experiments. 300GHz $\leq f \leq$ 3000GHz, and the inter-satellite distance is 100km $\leq d \leq$ 1000km. Both frequency and distance are uniformly sampled.

The transformer network parameters are shown in Table 1, and its learning parameters are shown in Table 2.

Table 1. Network parameters.

Network Layer	Parameters
Input layer	Input size: 66 Output size: 128
transformer	Input size: 128, Head: 8, Layer: 1
Output layer	Input size: 128, Output size: 11
Activation function	Relu

Table 2. Learning parameters.

Parameters	Value
Learning rate	0.0025
Batch size	1000
Risk factor	10% decays to 5% by epoch
Gamma	0.7
Max epochs	200

Sample 50, 200, and 400 data points from the free space path loss formula as the data set that the algorithm needs to fit. After the data set is fed into the algorithm model, the algorithm can automatically generate the best mathematical channel model.

3.4. Experimental Results

3.4.1. Convergence Effect

Figure 6 shows a typical training process, the blue curve is the average reward of all exploration results, and the red curve is the reward of the results used in training, that is, the average reward corresponding to the top 5–10% of the results, the orange curve is the optimal reward of each training epoch, and the black curve is the optimal reward of all training processes. It suggests that the average reward increases as the training continues. As can be seen from the black optimal curve, the best expression is explored in the 27th epoch, the reward reaches the highest value of 1, and it is confirmed that this is the optimal solution in subsequent epochs, and the convergence is completed. It reveals that the algorithm can quickly find the correct solution.

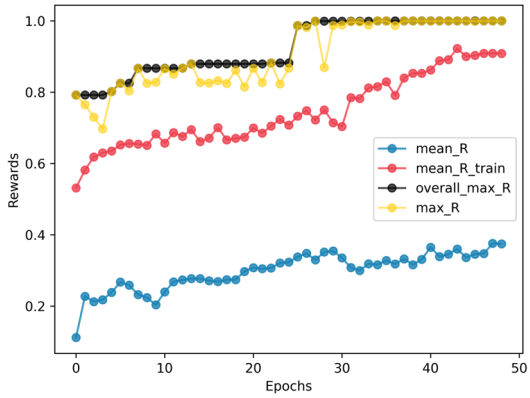


Figure 6. Reward convergence.

3.4.2. Fitting Effect

Figure 7 shows the fitting relationship between path loss and frequency. The black dots are sampled data points, the blue curve is all the explored results during training, the red curve is the best 5–10% of the results used for training, the orange curve is the best result of a training epoch, and the black curve is the best result of all epochs. One can see that the data points fall almost evenly on both sides of the black curve, which means that the black curve fits the distribution of the data points very well.

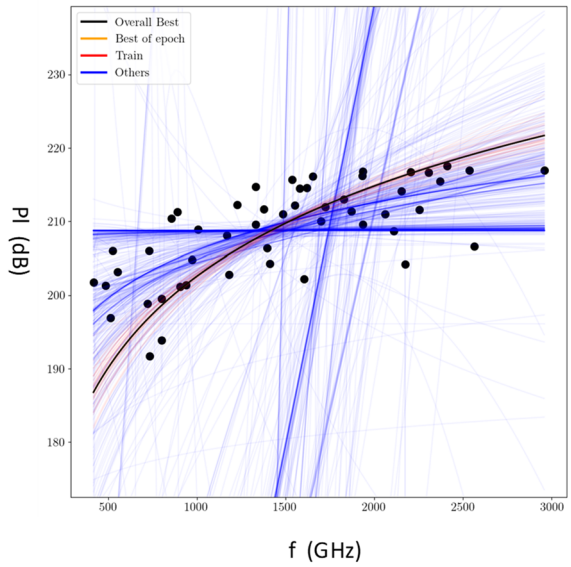


Figure 7. Relationship between path loss and frequency. The black dots are sampled data points.

Figure 8 depicts the relationship between the sampled data points and the curved surface, which is derived from the improved PhySO generated mathematical channel model $Pl = a\log(a^2)\log(a^2df) + a + b$, $a = 10.00, b = 42.44$, where a, b denote independent variables acting as coefficients in mathematical formulas. All the data points in red fall on the green curved surface, indicating that the improved PhySO finally fits the free space path loss correctly, and the improved algorithm has a good fitting effect on the inter-satellite terahertz channel model.

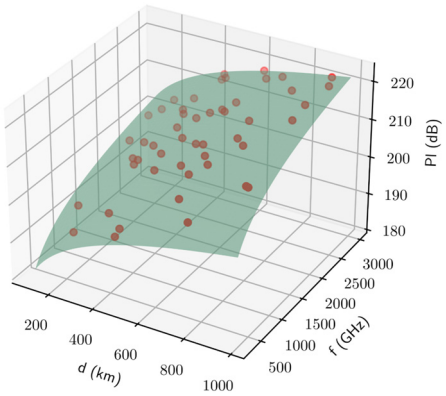


Figure 8. The surface of the generated model.

Table 3 depicts some typical free space path loss formulas generated by the improved PhySO. For instance, with 400 data points, the algorithm converges to expression $Pl = b\log(b\sqrt{adf})$, $a = 26.17, b = 40.00$. Post manual simplification, $Pl = 92.44 + 20\log(df)$. This formula corresponds to the free space path loss formula, validating the algorithm's precision in deducing the accurate formula for free space path loss from the communication dataset via symbolic regression.

Table 3. Typical Channel Expressions.

Number of Sampling Points	Expressions	(a, b)
50	$Pl = a\log(a^2)\log(a^2df) + a + b$	(10.00, 42.44)
200	$Pl = \sqrt{a} + \frac{a\log(adf)}{2} + 2b$	(40.00, 27.03)
400	$Pl = b\log(b\sqrt{adf})$	(26.17, 40.00)

3.4.3. Epochs Effect

Figure 9 shows the average training epochs required for generating the channel model of the LSTM network and the improved transformer network at 50, 200, and 400 data points. It can be seen that the transformer architecture requires fewer training epochs than the LSTM-based architecture, and fits the free space path loss model relatively faster.

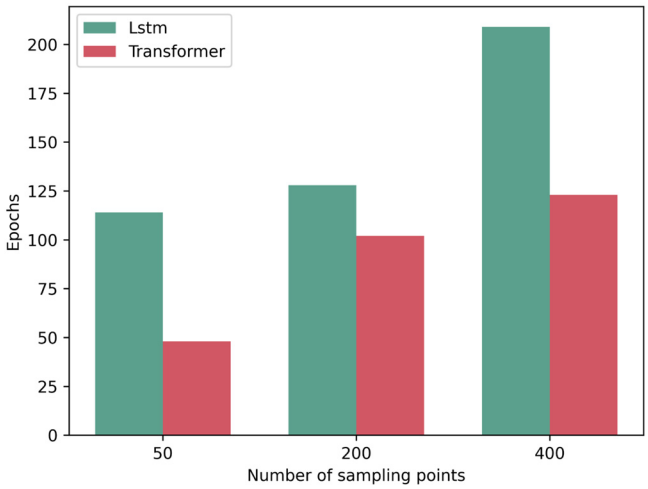


Figure 9. Relationship between sampling points and training epochs.

4. Conclusions

Channel modeling is a crucial technology in wireless communication. Traditional channel modeling methods are time-consuming, laborious, and require manual fitting of data. The scheme using a neural network directly as a channel model has the problem of interpretability, which makes it difficult to provide more insight into channels for researchers. The channel modeling method proposed in this paper based on symbolic regression uses a neural network as a tool to automatically generate a channel model, which not only saves labor costs but also avoids the "black box" characteristic of neural networks channel model. As a new channel modeling method, this paper successfully generates the free space path loss model based on the terahertz band from simulation data using the improved PhySO symbolic regression method and verifies the effect of symbolic regression on channel modeling. When the channel data is collected, the channel formula can be automatically obtained from the data by the proposed method, that is why the proposed method has an important application prospect in the field of channel modeling.

Of course, the project is not perfect, it still has problems such as the failure to get the final mathematical expression at one time, the unstable training epochs required for different random samples, and the need to retrain the network to establish the channel model each time facing new data sets. Therefore, further research will be conducted on the following topics:

1. To establish a usable inter-satellite terahertz channel model, it is necessary to use actual measured channel data for modeling by the proposed method in the future.
2. To expand the practicality of the proposed method, the channel modeling research will be further carried out from the aspects of path loss, delay power distribution,

multipath angle spatial distribution, Doppler frequency shift, and other aspects of terahertz communication.
3. To provide more concise and unified expressions, more efficient neural networks and deep reinforcement learning algorithms will be studied.
4. To reduce algorithm training times and improve symbol regression performance when using new data sets, the method based on pre-trained models is a worthwhile research direction.

Author Contributions: Conceptualization, Y.H., B.S. and Z.L.; methodology, B.S.; software, B.S.; validation, B.S. and Z.L.; formal analysis, B.S. and Z.L.; investigation, B.S. and Z.L.; resources, Y.H.; data curation, B.S. and Z.L.; writing—original draft preparation, Y.H., B.S. and Z.L.; writing—review and editing, Y.H., B.S. and Z.L.; visualization, B.S.; supervision, Y.H.; project administration, Y.H.; funding acquisition, Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Key Research and Development (R&D) Program, grant number 2019YFB1803201.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article.

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

References

1. Tian, H.; Tang, P.; Zhang, J. A review of terahertz channel characteristics and modeling for 6G. *Mob. Commun.* **2020**, *44*, 29–35+43.
2. Han, C.; Wang, Y.; Li, Y.; Chen, Y.; Abbasi, N.A.; Kürner, T.; Molisch, A.F. Terahertz wireless channels: A holistic survey on measurement, modeling, and analysis. *IEEE Commun. Surv. Tutor.* **2022**, *24*, 1670–1707. [CrossRef]
3. Walidainy, H.; Adriman, R.; Away, Y.; Nasaruddin, N. Channel modeling for 6G communications: A survey. In Proceedings of the 2021 International Conference on Computer System, Information Technology, and Electrical Engineering (COSITE), Banda Aceh, Indonesia, 20–21 October 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–6.
4. Serghiou, D.; Khalily, M.; Brown, T.W.; Tafazolli, R. Terahertz channel propagation phenomena, measurement techniques and modeling for 6G wireless communication applications: A survey, open challenges and future research directions. *IEEE Commun. Surv. Tutor.* **2022**, *24*, 1957–1996. [CrossRef]
5. Priebe, S.; Kurner, T. Stochastic modeling of THz indoor radio channels. *IEEE Trans. Wirel. Commun.* **2013**, *12*, 4445–4455. [CrossRef]
6. He, D.; Guan, K.; Ai, B.; Fricke, A.; He, R.; Zhong, Z.; Kasamatsu, A.; Hosako, I.; Kürner, T. Channel Modeling for Kiosk Downloading Communication System at 300 GHz. In Proceedings of the 2017 11th European Conference on Antennas and Propagation (EUCAP), Paris, France, 19–24 March 2017; pp. 1331–1335.
7. Bai, L.; Wang, C.X.; Huang, J.; Xu, Q.; Yang, Y.; Goussetis, G.; Sun, J.; Zhang, W. Predicting wireless mmWave massive MIMO channel characteristics using machine learning algorithms. *Wirel. Commun. Mob. Comput.* **2018**, *2018*, 9783863. [CrossRef]
8. Ferreira, G.P.; Matos, L.J.; Silva, J.M. Improvement of outdoor signal strength prediction in UHF band by artificial neural network. *IEEE Trans. Antennas Propag.* **2016**, *64*, 5404–5410. [CrossRef]
9. Xue, P.; Zhao, Y. A self-learning channel modeling approach based on explainable neural network. *IEEE Wirel. Commun. Lett.* **2023**, *12*, 1289–1293. [CrossRef]
10. Lee, H. Channel metamodeling for explainable data-driven channel model. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 2678–2682. [CrossRef]
11. Tenachi, W.; Ibata, R.; Diakogiannis, F.I. Deep symbolic regression for physics guided by units constraints: Toward the automated discovery of physical laws. *arXiv* **2023**, arXiv:2303.03192. [CrossRef]
12. Keren, L.S.; Liberzon, A.; Lazebnik, T. A computational framework for physics-informed symbolic regression with straightforward integration of domain knowledge. *Sci. Rep.* **2023**, *13*, 1249. [CrossRef] [PubMed]
13. Lu, Q.; Zhang, Y. Solving symbol regression based on Monte Carlo tree search. *Comput. Eng. Des.* **2020**, *41*, 2158–2164.
14. Petersen, B.K.; Landajuela, M.; Mundhenk, T.N.; Santiago, C.P.; Kim, S.K.; Kim, J.T. Deep symbolic regression: Recovering mathematical expressions from data via risk-seeking policy gradients. *arXiv* **2021**, arXiv:1912.04871.
15. Landajuela, M.; Petersen, B.K.; Kim, S.K.; Santiago, C.P.; Glatt, R.; Mundhenk, T.N.; Pettit, J.F.; Faissol, D.M. Improving exploration in policy gradient search: Application to symbolic optimization. *arXiv* **2021**, arXiv:2107.09158. [CrossRef]

16. 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. [CrossRef]
17. Moldovan, A.; Ruder, M.A.; Akyildiz, I.F.; Gerstacker, W.H. LOS and NLOS channel modeling for terahertz wireless communication with scattered rays. In Proceedings of the 2014 IEEE Globecom Workshops (GC Wkshps), Austin, TX, USA, 8–12 December 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 388–392.

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.