A FUZZY NEURAL NETWORK AND ITS GRADIENT-DESCENT ALGORITHM FOR PREDICTION INTERVALS

Nguyen Trong Trung Anh

Posts and Telecommunications Institute of Technology, Hanoi, Vietnam

Abstract: This study aims to propose a solution to handle the uncertainty and imprecise knowledge associated with the collected data using Interval type-2 fuzzy inference system, or IT2FIS. IT2FIS has been shown to be capable of generalizing functional relationship between input and output while reducing computational complexity. The proposed IT2FIS is a fuzzy neural network realizing Takagi-Sugeno-Kang inference mechanism. IT2FIS structure consists of multiple layers, which evolves automatically based on the incoming data. The parameters are updated using meta-cognitive learning and gradient descent algorithm. Prediction intervals are considered as the end-result of the system. For performance evaluation studies, collected data of wind speed and direction are utilized. Using historical data, the proposed model provides short-term forecasting of wind energy parameters. The performance of IT2FIS is compared with existing state-of-the art fuzzy inference system approaches and results clearly indicate the advantages of IT2FIS-based prediction.

Keywords: Neural network, interval type-2 fuzzy systems, meta-cognitive learning, gradient descent prediction intervals.

I. INTRODUCTION

Renewable energy gradually become more acceptable as an alternative source of energy due to its advantages over traditional fossil fuels, which in turn increases the complexity of the power system. One of the main reasons for the integration complexity is inaccurate prediction of these sources at a particular time. There are various natural issues as well as artificial causes. As a result, the forecast of generated energy is extremely uncertain. This uncertainty leads to unpredictable or unrealistic power generation, further leads to financial losses. Hence, realistic forecast of these sources is the need for increased and improved renewable energy usage. Recently, artificial neural network has been used in forecasting problems including solar energy forecasting, wind forecasting and significant wave height prediction [3] [4] [5]. It has been shown in the literature that neural network could be developed with fuzzy inference systems to enhance the learning capability. These types of systems are referred to as neuro-fuzzy inference system [11][12]. Traditionally, neuro-fuzzy inference systems employ type-1 fuzzy set, which has certain membership functions. In case of faulty data collection as invalid measurement, sensors error, or inaccurate data, the secondary membership of the type-2 fuzzy system provides a better resolution to handle uncertainties. Zadeh has proposed Type-2 fuzzy sets in [6] and Type-2 fuzzy logic systems employing these Type-2 fuzzy sets have been studied in various studies, and they are able to deal with uncertainties. However, the computation in type-2 fuzzy systems are massive. As a result, Interval Type-2 Fuzzy sets were proposed to minimize the computational effort in handling uncertainties associated with the data. Based on these Interval Type-2 fuzzy sets, different fuzzy inference systems, known as Interval Type-2 fuzzy inference system or IT2FIS, and their learning algorithm have been proposed in the literature [7] [2] [8]. Due to their fixed structures, research have shown that these algorithms were ineffective in dealing with the temporally varying data especially in a practical problem such as renewable energy forecasting. Meta-cognition is known as the fundamental ability to make decisions based on the surrounding environment. It is whether to learn a specific knowledge by using suitable learning strategies. In the literature, various studies on meta-cognitive algorithm integrated with neuro-fuzzy inference system have clearly shown their generalization abilities [9] [7]. The systems employing meta-cognitive learning are able to evolve theirs structures over training period. In the beginning, the structures are initialized with one rule, and they gradually change to adapt themselves according to incoming data.

The interval type-2 fuzzy neural network generating prediction intervals implemented in this study is based on gradient-descent learning algorithm. The algorithm named McIT2FIS-GD has been employed for several benchmark problems in the literature [1]. In this work, we extends the algorithm for prediction intervals in the wind energy domain and it is called FNN-PIs. The learning mechanism of FNN-PIs is formulated on a five-layer network realizing Takagi-Sugeno-Kang fuzzy inference. The input layer has n nodes, followed by K nodes of the membership layer. The firing

Contact author: Anh Nguyen Trong Trung

Email: anhntt@ptit.edu.vn

Manuscript received: 10/2021, revised: 11/2021, accepted: 12/2021.

layer calculates the firing strengths of K rules. The interval reduction layer uses the two q factors referred in [13] to construct output in the output layer. Meta-cognitive learning forms the learning mechanism of FNN-PIs. When an input comes, meta- cognitive component decides whether to delete the sample, learn the sample or reserve the sample based on the containing knowledge. By monitoring the novelty in arriving data, the system evolves over training process. The main contribution of this study is the learning algorithm of a first order optimization approach, gradient descent based fuzzy neural network. Additionally, the proposed algorithm is utilize to address the problem of prediction intervals compared to ordinary crisp target value.

In the next section, the architecture of the interval type-2 neural fuzzy inference system is presented. Meta-cognitive learning and gradient-descent based learning algorithm are deployed followed by performance evaluation on a real-world wind speed forecasting problem. The network is applied for prediction intervals and compared with other type-1 and type-2 systems in the literature.

II. META-COGNITIVE INTERVAL TYPE-2 NEURO FUZZY INFERENCE SYSTEM

Figure 1 illustrates the architecture of proposed meta-cognitive fuzzy neural network for prediction intervals, FNN-PIs. The structure consists of five layers. The aim is to estimate the non-linear functional relationship between *n-dimentional* input space and the output.

Layer 1 - Input layer: This layer contains n nodes representing n number of input features. The input is passed directly to layer 2, Membership Function Layer. The output of *i-th* node is given:

$$u_i(t) = x_i(t);\ i = 1, 2, \dots m. \quad (1)$$

Layer 2 - Fuzzification layer: This layer calculates the membership strength of each feature in every rule. Theformulas are given by:

$$\mu_{ij}^{up}(x_j) = \begin{cases} \varphi(m_{j1}^i, \sigma_j^i, x_j) & x_j < m_{j1}^i \\ 1 \quad if & m_{j1}^i \le x_j \le m_{j2}^i \\ \varphi(m_{j2}^i, \sigma_j^i, x_j) & x_j < m_{j2}^i \end{cases} \quad (2)$$

$$\mu_{ij}^{lo}(x_j) = \begin{cases} \varphi(m_{j2}^i, \sigma_j^i, x_j) & x_j \le \frac{m_{j1}^i + m_{j2}^i}{2} \\ \quad if & \\ \varphi(m_{j1}^i, \sigma_j^i, x_j) & x_j > \frac{m_{j1}^i + m_{j2}^i}{2} \end{cases} \quad (3)$$

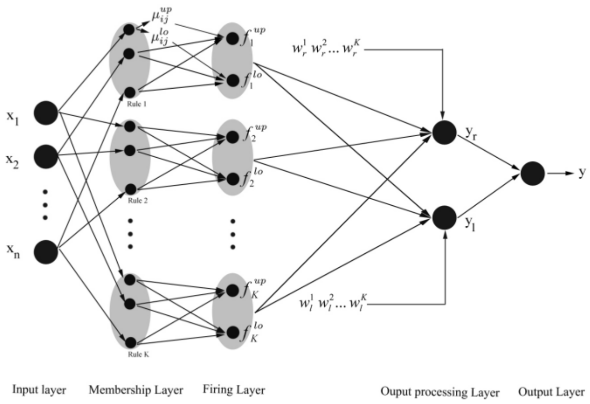

Figure 1. Architecture of the interval type-2 neuro fuzzy system

where,

$$\varphi(m_j^i, \sigma_j^i, x_j) = \exp\left(-\frac{(x_j - m_j^i)^2}{2(\sigma_j^i)^2}\right) \quad (4)$$

and m_{j1}, m_{j2}, σ are center left, center right and the width of the interval type-2 Gaussian membership function accordingly.

Layer 3- Firing layer: The firing strength of each rule iscalculated in this layer based on the following formulas:

$$f_i^{lo} = \prod_{j=1}^{n} \mu_{ij}^{lo} \quad (5)$$

$$f_i^{up} = \prod_{j=1}^{n} \mu_{ij}^{up} \quad (6)$$

Layer 4- Output processing layer: This layer has K nodes, each of which represents a rule in the network. Instead ofusing Karnick-Mendel iterative procedure to find lower and upper end points, this study uses q factors to increase the learning speed and minimize the computational complexity. The two end points are given as:

$$y_l = \frac{(1 - q_l)\sum_{i=1}^{K} f_i^{up} w_l^i + q_l \sum_{i=1}^{K} f_i^{lo} w_l^i}{\sum_{i=1}^{K}(f_i^{lo} + f_i^{up})} \quad (7)$$

$$y_r = \frac{(1 - q_r)\sum_{i=1}^{K} f_i^{lo} w_r^i + q_r \sum_{i=1}^{K} f_i^{up} w_r^i}{\sum_{i=1}^{K}(f_i^{lo} + f_i^{up})} \quad (8)$$

Layer 5- Output layer: The output of this layer is the sumof the two end points from the previous layer. The formula for estimated output at a certain training time period t is given as:

$$\hat{y}(t) = y_l(t) + y_r(t)\,. \quad (9)$$

III. META-COGNITIVE GRADIENT-DESCENT BASED LEARNING ALGORITHM FOR IT2FIS

- *Problem Definition*

The objective of proposed learning algorithm is to estimate the functional relationship ***f[.]*** so that the estimated output $\hat{y}$ is as close as the actual output y.

$$\hat{y}(t) = \boldsymbol{f}[x(t), \quad \boldsymbol{\theta}] \tag{10}$$

where, $\boldsymbol{\theta}$ describes the parameter vector of the neural architecture. In order to measure the novelty of the current sample, prediction error and spherical potential are considered. The error at the t-th given time is:

$$E(t) = \frac{1}{2}[y(t) - \hat{y}(t)] \tag{11}$$

The spherical potential is calculated as:

$$\psi(t) = \sum_{i=1}^{K} \frac{f_i^{lo}(t) + f_i^{up}(t)}{2K} \tag{12}$$

- *IT2FIS Gradient-descent learning algorithm*

A sample can be learn either by adding a rule or updating the parameters based on the containing knowledge in the sample. These criteria are known as Rule Adding Criteria and Rule Updating Algorithm respectively. The new rule is added into the system if the sample has significant knowledge and the existing rules are unable to cover the new sample effectively. Thus, we are going to define the rule addition criteria and the rule updating criteria to check against the prediction error presented by the contributing rules as well as the novelty of the sample.

A. The Rule Adding Criteria.

A new rule is grown if for the current sample, the predicted output is different from the desired output to a confidence level. In other words, the prediction error for the sample is higher than a threshold, and the novelty criterion is satisfied. The rule growing is based on the following condition:

$$E(t) > E_a \quad AND \quad \psi(t) < \psi_a \ . \tag{13}$$

where, E_a is the adding threshold and ψ_a is the novelty threshold. A higher value of E_a and a low value of ψ_a indicates resistance to rule addition. For problem chosen in this work, E_a and ψ_a were selected using grid search with five fold cross-validation.

B. The Rule Updating Algorithm.

The network parameters are updated if the prediction error of the current sample is higher than a parameters update threshold. The rule updating criteria is as follow:

$$E_u < E(t) < E_a \tag{14}$$

where, E_u denotes the self-adaptive parameter update threshold and decides if the parameters of the rules are updated using the gradient-descent based learning algorithm. The formulas for parameters updating in the network are given as:

$$m_{j1}^{i}(t+1) = m_{j1}^{i}(t) - \eta \frac{\partial E(t)}{\partial m_{j1}^{i}} \tag{15}$$

$$m_{j2}^{i}(t+1) = m_{j2}^{i}(t) - \eta \frac{\partial E(t)}{\partial m_{j2}^{i}} \tag{16}$$

$$\sigma^{i}\ (t+1) = \sigma^{i}\ (t) - \eta \frac{\partial E(t)}{\sigma^{i}} \tag{17}$$

$$w_l^{i}(t+1) = w_l^{i}(t) - \eta \frac{\partial E(t)}{\partial w_l^{i}} \tag{18}$$

$$w_r^{i}(t+1) = w_r^{i}(t) - \eta \frac{\partial E(t)}{\partial w_r^{i}} \tag{19}$$

where, η is the learning rate. The time dependence t indicates the current training iteration. The objective is to minimize the error function $E(t)$ defined in (11). The learning rate was chosen with care due to the impact of high learning rate. The problems were unable to converge if too high learning rates were employed while too low learning rates might lead to slow learning convergence.

Figure 2. Meta-cognitive learning mechanism

A self-regulatory learning algorithm forms the meta-cognitive learning mechanism of the system. During training process, the meta-cognitive component monitors knowledge presenting in the current sample to control appropriate learning strategy. The samples considered to be significant to the system are utilized to either grow a new rule and adapt the FNN structure or update the current existing fuzzy rules. The mechanism is illustrated in Figure 2. When a new sample is added into the network, the prediction error shows the knowledge difference between that particular sample and the inference of existing structure. Another measure to be taken into account is spherical potential as a novelty criterion [9].

IV. PERFORMANCE EVALUATION AND DISCUSSIONS

The performance evaluation of proposed interval type-2 fuzzy inference system and its gradient-descent learning algorithm is demonstrated on the benchmark data set of a non-linear system identification problem [15] and he real-world forecasting problem on the collected wind speed and direction. The performance is compared against type-1 neuro-fuzzy inference system including eTS, SAFIS [17], support vector regression (SVR) and type-2 neuro-fuzzy inference system including SIT2FNN [18], McIT2FIS [8]. The performance of the algorithm was evaluated in Matlab R2013b environment on a Windows system with Xeon CPU and 16GB RAM.

- *Performance Measures:*

In this study, root mean squared error is employed to evaluate the performance of the systems. Root mean squared error (RMSE) measures the difference between the actual and the predicted output and is defined as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y(t) - \hat{y}(t))^2} \qquad (20)$$

where, N indicates the total number of samples.

In order to evaluate the performance of the prediction intervals, coverage probability is to indicate the percentage of samples covered by the upper and lower bounds of the intervals. Prediction intervals coverage probability (PICP) used in [16] is defined as:

$$PICP = \frac{1}{N}\sum_{i}^{N} \in_i \qquad (21)$$

where, $\in_i$ shows whether the i-th sample was covered by the two bounds.

- *Wind Speed Prediction Problem:*

Recently, wind energy has emerged as a perennial source of energy due to a need for clean energy sources. The power generated by the wind turbines is nonscheduled in nature due to changing weather conditions. Since the prediction problem can be treated as a function approximation problem, FNN-PIs is employed for the wind prediction problem. The real-world wind data set is obtained from Iowa (USA) Department of Transport at the location, Washington. The data sampled for every 10 minutes is downloaded between over a period of one month. The data is averaged hourly from which ten features are extracted. The training date set consists of five hundred samples and the testing set consists of one hundred samples. Figure 3 shows the actual output versus the predicted output for training data. It can be observed from the figure that the underlying trend as well as functional relationship between the past-future wind speed was well-generalize. The performance of FNN-PIs was compared in this study. Table 1 shows the root mean square error for the benchmark algorithms in this study.

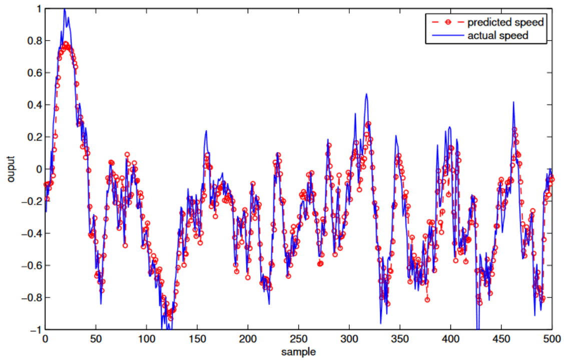

Fig. 3. Predicted and actual output for the training dataset using proposed interval type-2 neuro fuzzy system.

Algorithms employing type-1 and type-2 fuzzy sets are utilized. In order to predict the wind speed, it can be observed that FNN-PIs needs a total of six rules compared to thirty rules of FLANN. The training and testing root mean squared resulted from the proposed system are relatively low even though they are slightly higher than those of other type-2 meta-cognitive type-2 fuzzy neural networks in the literature. In order to study the learning progress of the proposed algorithms, We studied the impact of iterations over training. The logarithmic prediction error of the system during training process is demonstrated in Figure 4.

Type	Algorithm	Rules	RMSE	
			Train	Test
Type-1	SVR	-	4.026	4.363
	FLANN	30	-	0.532
Type-2	SIT2FNN	4	0.194	0.187
	McIT2FIS	4	0.136	0.132
	PBL-McIT2FIS	4	0.11	0.161
	FNN-PIs	6	0.162	0.153

Table 1. Performance evaluation on wind speed prediction problem.

It can be noticed from the figure that the prediction error decreases significantly within the first four hundred epochs. This can be explained by the fact that gradient descent algorithm optimized the network parameters based on the prediction error alone. In order to obtain a comprehensive prediction, it is experimented that the number of iterations varied in the range [800, 1000] for benchmark problem in this research.

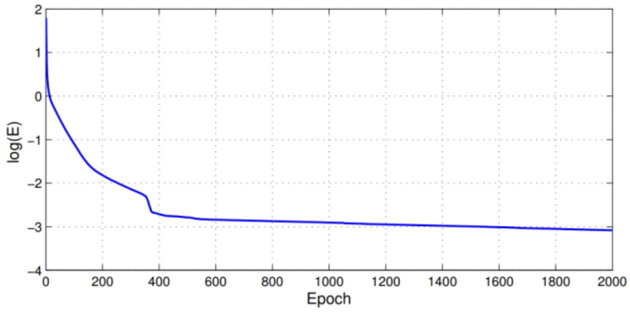

Fig. 4. Training logarithmic prediction error for benchmark problem using FNN-PIs

In term of prediction intervals, this research was conducted on the same wind data set. Coverage probability and average width are measured to track the performance of FNN-PIs. Prediction intervals coverage probability has been described in (21), and the average width of the intervals, PIAW [16] is given as:

$$PIAW = \frac{1}{N}\sum_{i=1}^{N}\frac{(U_i - L_i)}{R} \quad (22)$$

where L_i and U_i are lower and upper bound, respectively. R is the range of the underlying targets. Table 2 shows the evaluation indicators for training, testing data and the percentage of samples employed in two IT2FISs in this studies. From the table we are able to notice that FNN-PIs quantifies a proper prediction for the wind speed data while learning 65.2% of the samples.

Table 2. Prediction Intervals evaluation for training and testing set

		Training	**Testing**
RMcIT2FIS	PICP	0.81	0.92
	PIAW	0.24	0.35
		**PS = 67.8%*	
FNN-PIs	PICP	0.92	0.84
	PIAW	0.31	0.39
		**PS = 65.2%*	

V. CONCLUSIONS

In this study, an interval type-2 fuzzy inference system with its meta-cognitive learning algorithm is evaluated on the wave rider data set. Meta-cognitive component assesses the knowledge presenting in each sample to decide what-to- learn, how-to-learn and when-to learn effectively. The sample is learn in case the knowledge present is novel to the system while it is deleted when the same knowledge has been there. Learning strategies are based on the novelty criterion as spherical potential and energy function criterion as hinge loss error. Based on this tactics, it is ensure that the redundant computation is avoided and the algorithm runs efficiently. The performance is evaluated on a significant wave height prediction problem and it has been shown that the algorithm can generalize the trend well.

REFERENCES

[1] A. K. Das, N. Anh, S. Suresh, and N. Srikanth, "An interval type-2 fuzzy inference system and its meta-cognitive learning algorithm," *Evolving Systems*, vol. 7, no. 2, pp. 95–105, 2016. [Online]. Available: http://dx.doi.org/10.1007/s12530-016-9148-6

[2] Y. Y. Lin, S. H. Liao, J. Y. Chang, and C. T. Lin, "Simplified interval type-2 fuzzy neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 5, pp. 959–969, May 2014.

[3] A. K. Das, S. Suresh, and N. Srikanth, "Meta-cognitive interval type- 2 neuro-fuzzy inference system for wind prediction," in *Multisensor Fusion and Information Integration for Intelligent Systems (MFI), 2014 International Conference on*, Sept 2014, pp. 1–6.

[4] M. zger, "Significant wave height forecasting using wavelet fuzzy logic approach," *Ocean Engineering*, vol. 37, no. 16, pp. 1443 – 1451, 2010. [Online]. Available:

[5] M. A. Hashim, J. Jaafar, and S. M. Taib, "Interval type 2 fuzzy logic for significant wave height forecasting," in *2012 12th International Conference on Intelligent Systems Design and Applications (ISDA)*, Nov 2012, pp. 794–798.

[6] Q. Liang and J. M. Mendel, "Interval type-2 fuzzy logic systems: theory and design," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 5, pp. 535–550, Oct 2000.

[7] A. K. Das, K. Subramanian, and S. Suresh, "A computationally fast interval type-2 neuro-fuzzy inference system and its meta-cognitive projection based learning algorithm," in *2014 International Joint Con- ference on Neural Networks (IJCNN)*, July 2014, pp. 1510–1516.

[8] A. K. Das, K. Subramanian, and S. Sundaram, "An evolving interval type-2 neurofuzzy inference system and its metacognitive sequential learning algorithm," *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 6, pp. 2080–2093, Dec 2015.

[9] K. Subramanian and S. Suresh, "A meta-cognitive sequential learning algorithm for neuro-fuzzy inference system," *Applied Soft Computing*, vol. 12, no. 11, pp. 3603 – 3614, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1568494612002888

[10] J. S. R. Jang, "Anfis: adaptive-network-based fuzzy inference system," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 23, no. 3, pp. 665–685, May 1993.

[11] Hideyuki Takagi, "Fusion technology of neural network and fuzzy system: a chronicled progession from the laboratory to our daily lives." 10,01 2000.

[12] Isao Hayashi and Motohide Umano, "Perspectives and trends of fuzzy neural networks". J*ournal of Japan Society for Fuzzy Theory and Systems*, 5(2):178-190, 1993.

[13] R. H. Abiyev and O. Kaynak, "Type-2 fuzzy neural network structure for identification and control of time-varying plants". *IEEE Transactions on Industrial Electronics*, 57(12):4147-4159, 2010.

[14] A. Khosravi, S. Nahavandi and D. Creighton, "Prediction interval construction and optimization for adaptive neuro fuzzy inference systems". *IEEE Transactions on Fuzzy Systems*, 19(5):983-988, 2011.

[15] Chia-Feng Juang and Chin-Teng Lin, "An online self-

constructing neural fuzzy inference network and its applications". IEEE Transactions on Fuzzy Systems, 6(1):12-32, 1998.

[16] Hao Quan, D. Srinivasan and A. Khosravi. "Construction of neural-network based prediction intervals using particle swarm optimization". *In the 2012 International Joint Conference on Neural Networks (IJCNN)*, pages 1-7, 2012.

[17] Sequential adaptive fuzzy inference system (safis) for non-linear system identification and prediction. *Fuzzy sets and Systems*, 157(9):1260 - 1275, 2006.

[18] Y. Y. Lin, S. H. Liao, J. Y. Chang, and C. T. Lin. *Simplified interval type-2 fuzzy neural networks. IEEE Transactions on Neural Networks and Learning Systems*, 25(5):959-969, May 2014.

MẠNG NƠ-RON LOGIC MỜ VÀ THUẬT TOÁN SUY GIẢM ĐỘ DỐC CHO DỰ ĐOÁN KHOẢNG

Tóm tắt: Nghiên cứu nhằm mục đích đề xuất giải pháp xử lý dữ liệu không chắc chắn thu thập được từ cảm biến bằng cách sử dụng hệ thống mạng nơ-ron logic mờ kiểu 2, hoặc IT2FIS. IT2FIS đã thể hiện được khả năng tổng quát hóa mối quan hệ chức năng giữa đầu vào và đầu ra trong khi giảm độ phức tạp tính toán. Mô hình IT2FIS đề xuất là một mạng nơ-ron logic mờ thực hiện cơ chế suy luận Takagi-Sugeno-Kang. Cấu trúc IT2FIS bao gồm nhiều lớp, tự động phát triển dựa trên dữ liệu đến. Các thông số được cập nhật bằng cách sử dụng phương pháp học có nhận thức và thuật toán giảm độ dốc. Các khoảng dự đoán là kết quả cuối cùng của thuật toán. Sử dụng dữ liệu lịch sử, mô hình được đề xuất cung cấp dự báo ngắn hạn về các thông số năng lượng gió.

Từ khóa: Mạng nơ-ron, học máy, logic mờ.

Nguyen Trong Trung Anh, nhận học vị Tiến sỹ năm 2018 tại đại học công nghệ Nanyang, Singapore. Hiện công tác tại Học viện Công nghệ Bưu chính Viễn thông. Lĩnh vực nghiên cứu: Trí tuệ nhân tạo, xử lý dữ liệu, học máy, mạng nơ-ron.

Email: anhntt@ptit.edu.vn