

APPLYING MACHINE LEARNING IN NETWORK TOPOLOGY OPTIMIZATION

Zhouwei Gang, Qianyin Rao, LinGuo, LinXi, Zezhong Feng, QianDeng
GuanShanXi Road, Guanshan Lake District, GuiYang GuiZhou, China

NOTE: Corresponding author: Zhouwei Gang (13595101082@139.com)

Abstract – Nowadays, telecommunications have become an indispensable part of our life, 5G technology brings better network speeds, helps the AR and VR industry, and connects everything. It will deeply change our society. Transmission is the vessel of telecommunications. While the vessel is not so healthy, some of them are overloaded, meanwhile, others still have lots of capacity. It not only affects the customer experience, but also affects the development of communication services because of a resources problem. A transmission network is composed of transmission nodes and links. So that the possible topology numbers equal to node number multiplied by number of links means it is impossible for humans to optimize. We use AI instead of humans for topology optimization. The AI optimization solution uses an ITU Machine Learning (ML) standard, Breadth-First Search (BFS) greedy algorithm and other mainstream algorithms to solve the problem. It saves a lot of money and human resources, and also hugely improves traffic absorption capacity. The author comes from the team named "No Boundaries". The team attend ITU AI/ML in 5G Challenge and won the Gold champions (1st place).

Keywords – 5G, artificial intelligence (AI), data handling, intelligence level, machine learning (ML)

1. BACKGROUND

Nowadays, telecommunications have become an indispensable part of people's lives. We connect to the Internet through mobile phones for social networking (Facebook, Weibo, etc.), entertainment (YouTube, mobile games, etc.), e-shopping (Amazon, eBay, etc.) and work (remote office OA, etc.), and the upcoming 5G network has expanded from people-to-people communication to people-to-things, and things-to-things communication. Therefore, the future network will be more deeply embedded in our society and life. Assuming that a telecommunication system is a human body, transmission is as important as a vessel system which sends our data (also known as traffic) to the place it is needed. To construct a highly efficient network, each node should be used efficiently to form the transmission network.

However, just as a large number of diseases are related to a blood vessel, the situation of the transmission network is not optimistic in the actual communication network.

Due to the characteristics of mobile communications, increasingly complex mobile Internet services, and increasing traffic, the utilization rate of the transmission network is generally unbalanced. Congestion is caused by the overload of a small number of key nodes, and about half of the nodes are under low load. Take the China

Mobile Kaili PTN network as an example. Among the 913 links in the entire network, high-load links account for 9% while low-load links reach 53%. Operations staff need to optimize the network topology.

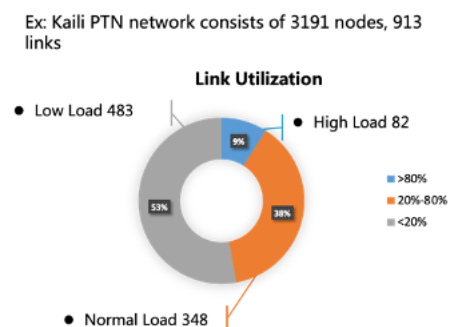


Fig. 1 – Link utilization

A transmission network is composed of transmission nodes and links. So that the possible topology numbers equal the node number multiplied by the number of links, taking Kaili as an example, the possibility of all topologies exceeds 2.9 million, which is impossible for humans to optimize.

So, at present, we can only adopt two methods, one is to adjust only one link locally. We optimize some of the chains with ultra-high utilization rates by experience or artificial prediction to reduce the number of nodes and load on a single link. And the other is to directly upgrade the hardware, which increases the bandwidth capacity of the entire link

by expanding hardware boards and ports to reduce link utilization (traffic/bandwidth). Due to the shortcomings and limitations of the current scheme, it is difficult to improve the current unbalanced transmission network.

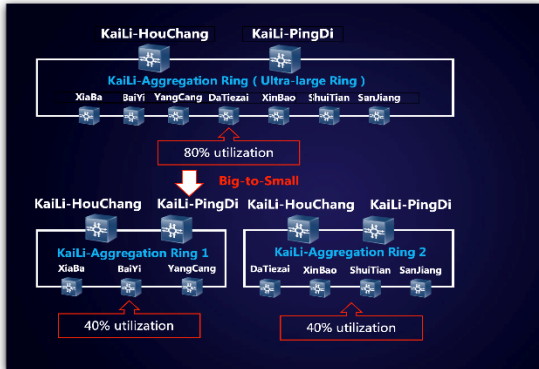


Fig. 2 – Solution1:Local(Ring) Adjustment: "big-to-small" optimization for some links, reducing the number of nodes and load

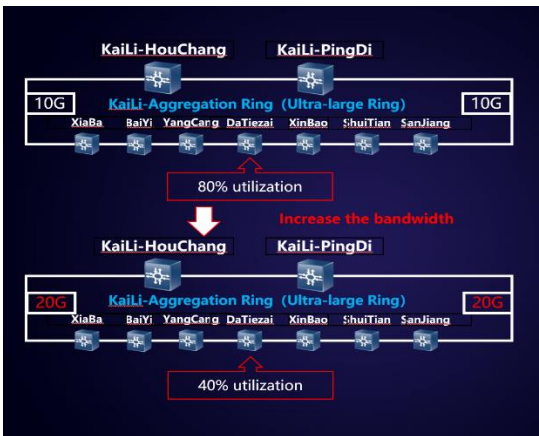


Fig. 3 – Solution2:Node expansion: Increase bandwidth of the link by upgrading hardware, and reduce link utilization

To solve this problem more effectively, the No Boundaries team has read a lot of literature, but there is no mature solution.

Qianyin Rao believes that applying the new SDN technology of 5G transmission network to transmission network can only solve the problems of delay and local optimization, and cannot support global topology optimization[1]. Intelligent control is still insufficient.

Lin Guo believes that wireless devices are the edges of networks, which differ from the nature of the transmission equipment[2] [3]. So a wireless network topology optimization scheme cannot be adopted.

Zhouwei Gang believes that the number of nodes in the data center and Internet scenarios is far less than that of the transmission network and cannot be applied[4] [5] [6].

Although no mature solution was found, the team was inspired to use machine learning instead of humans for topology optimization.

2. SOLUTION

2.1 Simulated environment

In order to explore the optimization of network topology based on machine learning, China Mobile selected more than 2,000 pieces of node topology information in three cities A, B, and C from the real network and the traffic situation of the nodes in the previous 20 days as the basic data of the contest, and the comprehensive evaluation indexes of topology optimization results (single day) are as follows:

$$\text{max object_ratio}_{\text{avg}} - (E_{\text{avg}} + E_{\text{min}} + E_{\text{max}}) - \alpha \cdot \text{sub_ratio} - \beta \cdot \text{hang_ratio}$$

$\text{object_ratio}_{\text{avg}}$: the hourly daily average value of the ratio of link bandwidth utilization optimized to the target range;

E_{avg} : hourly daily average value of the network topology load balancing E value;

E_{min} : hourly daily minimum value of the network topology load balancing E value;

E_{max} : hourly and daily maximum value of the network topology load balancing E value;

sub_ratio : ratio of the number of nodes on the secondary link in each link to the number of nodes in the entire link (hourly daily average);

hang_ratio : the ratio of the number of nodes hanging in each link to the number of nodes in the entire link (hourly daily average);

α, β : Constant coefficients ($\alpha=0.02, \beta =0.05$).

The specific algorithm of the evaluation index is as follows:

Target range of link bandwidth utilization:

$$\left[\frac{3}{5} \times \mu, \frac{7}{5} \times \mu \right]$$

The calculation method of the load balancing E value is as follows:

E value = variance of the fp value of all links in the network topology

$$\text{Link bandwidth utilization } f_p = \frac{\sum_i^{\text{nodes}} f_{X_i}}{A}$$

The μ in the upper and lower limits of the target range is the average value of bandwidth utilization of all links in the network topology.

In the f_P calculation formula, f_X is the flow value of the network element nodes in the link except the link head and tail nodes, and A is the maximum value of the A value of the other nodes in the link except the link head and tail nodes.

2.2 Architecture design

Through team analysis, Qian Deng found ITU's machine learning framework in the future network (mainly containing three components, ML sandbox system, ML pipeline subsystem and management subsystem), and believed that the ML pipeline subsystem met the needs of this competition.

The ML pipeline subsystem consists of 7 parts, but the data has been provided for this competition, and the optimization results are given in the form of a table and do not need to be directly connected to the equipment. Therefore, SRC, C, D, and SINK are not involved in the development. The result mainly consists of three parts: PP for data cleaning, M for topology restoration and traffic prediction [9], and P for optimization of the topology according to optimization rules and predicted traffic strategy.

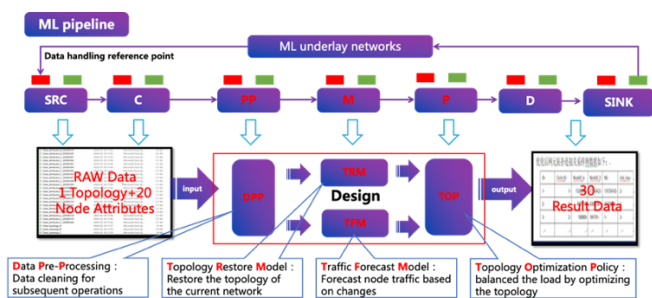


Fig. 4 - Mapping ITU architecture

2.3 Algorithm selection

Since the data comes from the real network, there are certain deficiencies, so the data preprocessing mainly considers two aspects of data integrity and ease of use:

Data integrity: Zezhong Feng uses pandas to check the integrity of key fields (traffic, latitude, longitude, connection relationship, etc.) and fill in missing data to ensure normal operation of subsequent predictions and optimizations.

Ease of use: Lin Xi believes that data is based on nodes. To restore the connection of nodes in the network topology, it is necessary to transform the data structure to facilitate subsequent calls.

Compared with the adjacency table, using an adjacency matrix to store the connection relationships of nodes can improve query efficiency.

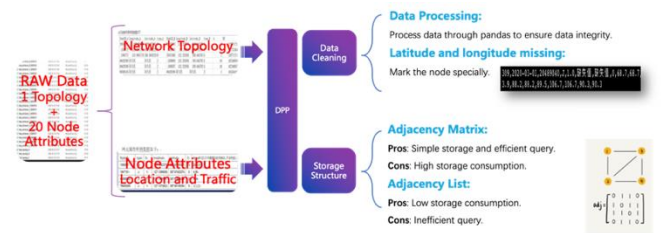


Fig. 5 - Node structure

2.4 Modeling

Regarding the Topology Restoration Model (TRM) and Traffic Forecast Model (TFM), Zhouwei Gang believes that the essence of topology restoration is to organize and form a new data set according to the specified conditions from the original data set. Therefore, search algorithms can be used for processing. Traffic forecasting is based on the changes of things in the past, mining the law of change, predicting the future, and there is a strong correlation with time, so the time series prediction algorithm is used for processing [10][11].

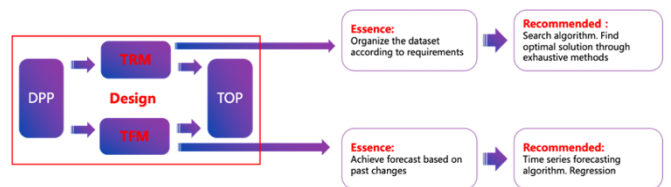


Fig. 6 - Time series prediction algorithm

Before TRM, Data Preprocessing (DPP) had converted the node connection relationship into a graph representation. Therefore, the process of finding a set of links in the topology that meets the specified conditions (topology restoration or topology optimization) is essentially a graph search problem. The most commonly used search algorithms for graphs are Breadth-First Search (BFS) and Depth-First Search (DFS) algorithms. Lin Xi said that this project has certain requirements for convergence time, so the breadth-first search algorithm is adopted for higher search efficiency.

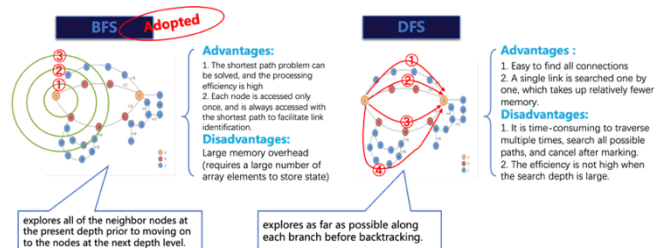


Fig. 7 - BFS and DFS

However, traditional BFS cannot visit the visited node, and there is a situation in which the node is visited multiple times in the simulation data. To solve this problem, Lin Xi modified the BFS search status part so that the improved algorithm can repeatedly visit the node and adapt the search of the link set.

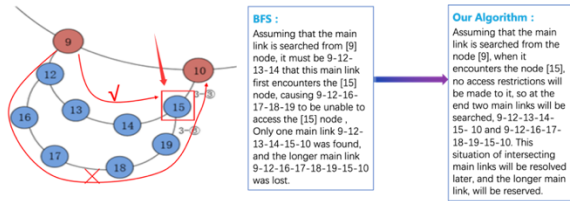


Fig. 8 – Improved BFS

For TFM, mainstream time series prediction algorithms include multiple linear regression, Autoregressive differential Moving Average (ARIMA) and Long-Short-Term Memory network (LSTM) model [7]. By analyzing the characteristics and data of this research, the influence between the data is almost non-linear. Among the above prediction models, the LSTM model is designed with a special structure to memorize and filter the changes in traffic on the time scale. Therefore, we choose the LSTM recurrent neural network model for prediction to meet the requirements of this prediction scenario.

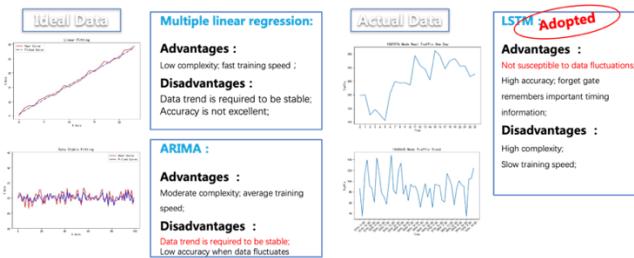


Fig. 9 – LSTM adopted

Ze Zhong Feng believes that the bottleneck for improving the accuracy of traffic forecasting is that there are few samples and few influencing factors, many nodes of different types, so the model is in two layers to reduce overfitting and under-fitting. We build a model for each node and increase the flow information carried by each neuron from 1 to 24, so that the model can fit more flow changes. In the end, the average accuracy of TFM increased from 91.7% to 95.3%.

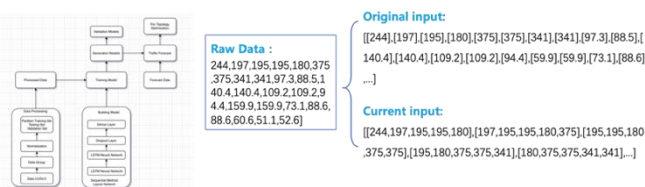


Fig. 10 – Improved input

In order to further improve the prediction accuracy, Zezhong Feng tried to add more features to improve the traffic forecast model. In the time dimension, we add the two features of the day's weather and the weekend to make it have traffic and weekend features; in the space dimension, we add the node traffic at a certain distance around the node as a feature, and we analyze that it is within a certain range (tentative 2 Km) where nodes have an impact on the node traffic of the current link, and such node traffic is added to features.

Ze Zhong Feng's experimental results indicate that:

By adding weather features, the average forecast accuracy of the model is increased by 0.4% ;

by adding the features of working day/off day, the average forecast accuracy of the model is increased by 1.6% ;

by adding the traffic chromatistics of surrounding nodes, the average forecast accuracy of the model is increased by 1.8% ;

after all three features were added, the average forecast accuracy of the model increased from 91.7% to 95.3%.

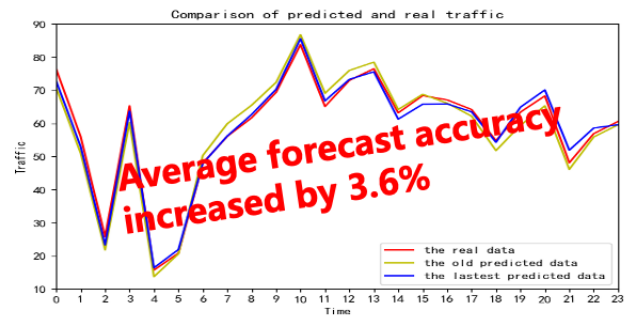


Fig. 11

2.5 Optimization strategy

In terms of Topology Optimization Strategy (TOP), Zhouwei Gang proposed an optimization strategy of "(ultra-low)*3". First calculate the utilization rate of all links and the average value of the entire network, and set a threshold to divide all links into three categories: overload, low load and normal. The overload links are processed first, and then the low-load links are processed. After finishing, adjust the threshold value, do it twice again, and finally obtain the optimized topology by completing three threshold adjustments.

1. Optimize overloaded links
2. Optimize low-load links
3. Adjust the threshold and perform step 1 and step 2 again
4. Perform step 3 again
5. Finish

threshold	overloaded	low-load
1	>2 μ	<0.2 μ
2	>1.6 μ	<0.4 μ
3	>1.4 μ	<0.6 μ

μ = Average link utilization of the entire network

Fig. 12 – Topology optimization strategy

Zhouwei Gang completed the optimization rules for overload and low-load links, and Lin Xi realized the aggregation of low-load links with the same head and tail nodes.

Overload link optimization scheme:

Calculate the utilization rate of all links and the average value of the entire network based on the predicted node traffic, search for all overloaded links to form a set, and select a link, as shown in the red link on the left. Along the link sequence, each node is judged according to rules, whether it can establish a link with other link nodes, and the results are formed into a set. According to the load of the original link and the load of the new link, find a pair of nodes from the set, disconnect them, and connect them to other links to form a new link.

The low-load link optimization rules are similar and will not be elaborated.

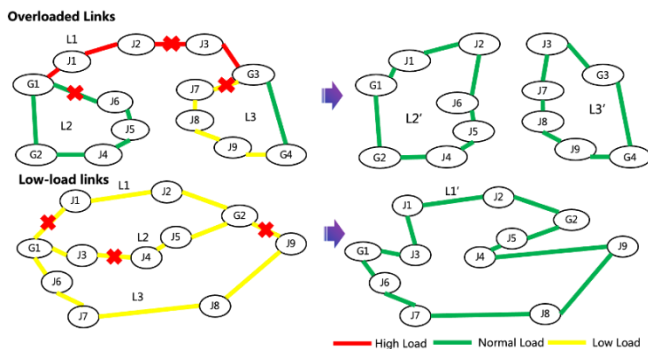


Fig. 13 – Link adjustment

L stands for link, J stands for J node. Through the "(ultra low)*3" TOP (optimization strategy refers to the calculation of three operations repeated three times):

1. Optimize overloaded links
2. Optimize low-load links
3. Adjust the threshold and recalculate the overload link and the low-load link

The algorithm can complete topology optimization in 55.5 minutes. Lin Guo used Qunee (Web graphics component solution) to complete the topology map. The left picture of the figure below is the original topology of City C, and the right picture is the optimized topology. Green nodes indicate adjusted nodes.

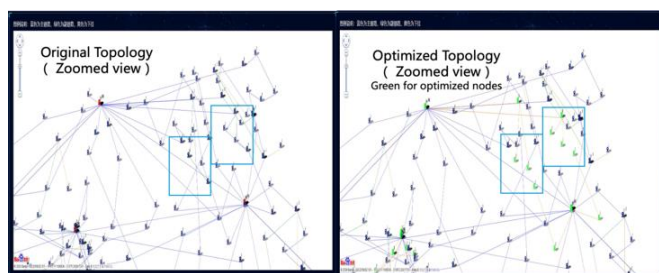


Fig. 14 – Node load reduction

But the "(ultra-low)*3" TOP (topology optimization strategy) is still the topology optimization of the link granularity, can it be further refined to the node granularity?

Lin Xi introduced a greedy algorithm to improve the optimization strategy. The core strategy is to refer to the index score when optimizing overloaded links and low-load links, and always move the nodes to get the highest index score. In this way, the node is moved to the optimal link.

The calculation of the index score is mainly related to the link optimization ratio, the average, minimum, and maximum value of link bandwidth utilization, and the ratio of secondary links and downstream nodes in the link. Using the optimization strategy of the greedy algorithm, the nodes are not randomly moved to other links, but selectively moved to the links that make the index score higher, and the optimization granularity is refined from the link level to the node level, thereby improving the overall optimization effect, with the link optimization ratio increased by 0.4. Although the calculation time doubled, it did not exceed 15 minutes, and the optimization evaluation score increased by about 10%.

At the same time, Lin Xi has tested the genetic algorithm. For the topic of network topology optimization, the genetic algorithm involves more debugging such as parameter coding, selection, crossover, mutation, etc. For example, topology optimization is difficult to map and transform in the parameter coding part, and it involves rule constraints during crossover, considering the entire genetic algorithm for topology optimization is actually very complicated, finally only the greedy algorithm is used.

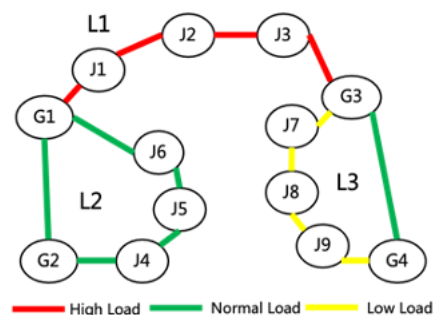


Fig. 15 – Greedy algorithm

	J1	J2	J3	J4
L1	0.4618	0.5026	0.5185	0.5531
L2	0.4767	0.5345	0.5567	0.5732
L3	0.5068	0.5123	0.5234	0.5613

The link optimization ratio increased 11%

Fig. 16 - Increased 11%

2.6 Real-world scenario

Qianyin Rao introduced the results into Kaili City, with 913 chains composed of 3191 network elements, and an AI optimization algorithm output of 60 high-load solutions, of which 29 have passed the expert review and are qualified for implementation. In order to ensure safety, the operation is carried out in batches. Each optimization does not exceed 50 nodes, and the scope of influence is controlled within 10 links. After the operation is completed, observe the stability of the network for 3 days before proceeding to the next operation. Up to now, 11 batches of adjustments have been completed, and 16 high-load links have restored reasonable loads.

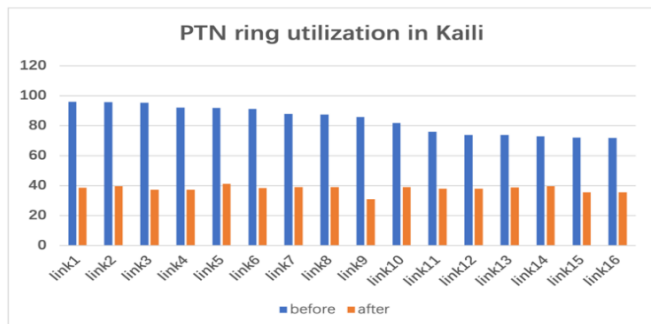


Fig. 17 - PTN ring utilization

For these 16 overloaded links, compared with the traditional solution, this network topology optimization saves construction costs by 2.6 million RMB, reduces personnel investment by 17 man-days, and increases network absorption traffic by 79TB/day.

Plan	Traditional Plan	AI Plan
Construction Cost	¥ 3.05M	¥ 0.45M
Human Input	21 Man-Days	4 Man-Days
Low-load links	40	3
High-load links	5	0
Capacity (day)	102T	181T

Fig. 18 - Economic performance

2.7 Standard packaging and rapid promotion

Lin Guo said, 'In order to meet the needs of rapid replication and promotion, according to ITU ML architecture and recommendations, the requirements in ITU recommendations are implemented by docker and Kubernetes'. Docker container is a standard unit of software that packages up code and all its dependencies so the application runs quickly and reliably from one computing environment to another K8s, is an open-source system for automating deployment, scaling, and management of containerized applications. According to the test, the development time is shortened by 70%, data sharing is applied to reduce 50% storage, and fault self-healing is realized. Finally, the overall architecture is completed on the basis of integration with ITU architecture.

3. RESULTS AND DISCUSSION

3.1. Results

In accordance with the ITU standard Lin Guo evaluated the results, the application results reached L3 in analysis, decision and demand mapping, and the data collection reached L2. Since the newly-built link in the transmission topology optimization involves physical entity operations and cannot be fully implemented by the software system, the action implementation only reaches L1.

But the team believes that after the results continue to improve, the three parts of data collection, analysis and decision can be upgraded from the current L2 and L3 levels to L5. The demand mapping team believes that it needs to accumulate more expert experience to break through to L4. If the results are applied to other similar network topology optimizations that do not require physical adjustment, the team believes that L5 can be achieved.

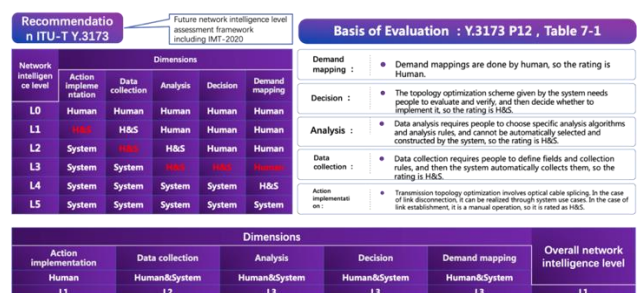


Fig. 19 - Intelligent level

At the same time, the team believes that there is still room for further evolution of this result.

First of all, continuing to optimize the algorithm is a good choice. At present, we only use the greedy algorithm, but there are actually many optimization algorithms that have not been tested and verified for similar problems.

Secondly, conduct pilot projects in more locations to accumulate sample data to adjust and optimize strategies to improve accuracy.

Finally, after the topology adjustment is completed, perform a certain period of observation, roll back, and analyze the reasons for those that have not reached expectations, and guide the improvement of the strategy.

3.2. Discussion

Through this research, with the support of ITU planning, our results have achieved breakthrough success. Compared with traditional operation and maintenance methods, we have obvious advantages. We think that this AI/ML application in the communications industry has positive significance in three aspects:

The communication network before 5G mainly communicates between people, and the three typical scenarios in the 5G standard reflect that 5G has increased the communication between people and things, and between things. This type of communication will not only change the way of human life, but also change the mode of social operations. Facing this great prospect, ITU, global operators and equipment vendors are all vigorously promoting the construction and operation of 5G networks. But this also causes the complexity of the 5G network to far exceed the previous network, and the construction cost and operation and maintenance cost have doubled. The results of this combination of communication operation and maintenance and AI can greatly reduce the cost of 5G network construction and operation and maintenance, and accelerate the promotion of 5G. Therefore, it is recommended that the ITU continues to improve the integration of AI/ML technologies in the communication specifications to provide support for global operators and equipment vendors, so as to burst more intelligent operation and maintenance results, promote 5G networks to change society as soon as possible, and enable us to enter a better future.



Fig. 20 – 5G for everything

In China Mobile, an intelligent operation and maintenance ecosystem of developers + platform + standards have been formed. The "Nine days" intelligent intermediate platform created by China Mobile provides functions for developers and gathers the wisdom of developers who work in accordance with ITU standards. The result is screened by China Mobile and then promoted, forming an endless ecosystem. Relying on this ecosystem, China Mobile has cultivated many intelligent operation and maintenance results, supports China Mobile in building the world's largest 5G network, and constantly discovers new problems in construction, operation and maintenance, and new problems have become new goals for developers. In this way, we can realize the efficient operation and maintenance of complex networks.



Fig. 21 – ecological environment

Most of the current AI development is the to customer mode that focuses on serving users, based on user needs, and conforms to users' living habits. Based on the ITU specifications, this project has achieved the integration of the communications industry and AI. This is an exploration of the to business model of AI development serving the industry. It is developed based on industry needs and meets industry development specifications. It is foreseeable that more and more AI services will adopt the industry-to-B model in the future, this research will provide an example for them.



Fig. 22 – AI to C and AI to B

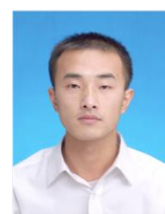
REFERENCES

- [1] Rusek, K., Suárez-Varela, J., Mestres, A., Barlet-Ros, P., & Cabellos-Aparicio, A. (2019, April). Unveiling the potential of Graph Neural Networks for network modeling and optimization in SDN. In Proceedings of the 2019 ACM Symposium on SDN Research (pp. 140-151).
- [2] Gui, J., & Zhou, K. (2016). Flexible adjustments between energy and capacity for topology control in heterogeneous wireless multi-hop networks. *Journal of Network and Systems Management*, 24(4), 789-812.
- [3] Ma, R., Cao, X., Zhang, S., Liu, L., Cheng, Y., & Sun, C. (2018, December). A DBN-Based Independent Set Learning Algorithm for Capacity Optimization in Wireless Networks. In 2018 IEEE Global Communications Conference (GLOBECOM) (pp. 1-6). IEEE.
- [4] Heller, B., Seetharaman, S., Mahadevan, P., Yiakoumis, Y., Sharma, P., Banerjee, S., & McKeown, N. (2010, April). Elastictree: Saving energy in data center networks. In *Nsdi* (Vol. 10, pp. 249-264).
- [5] Zhuo, D., Ghobadi, M., Mahajan, R., Förster, K. T., Krishnamurthy, A., & Anderson, T. (2017, August). Understanding and mitigating packet corruption in data center networks. In Proceedings of the Conference of the ACM Special Interest Group on Data Communication (pp. 362-375).
- [6] Perozzi, B., Al-Rfou, R., & Skiena, S. (2014, August). Deepwalk: Online learning of social representations. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 701-710).
- [7] F.A. Gers, J. Schmidhuber and F. Cummins, Learning to forget: Continual prediction with lstm, 1999.
- [8] Qing He, Arash Moayyedi, György Dán, Georgios P. Koudouridis, Per Tengkvist, "A Meta-Learning Scheme for Adaptive Short-Term Network Traffic Prediction", *Selected Areas in Communications IEEE Journal on*, vol. 38, no. 10, pp. 2271-2283, 2020.
- [9] Y. Tian and L. Pan, "Predicting Short-Term Traffic Flow by Long Short-Term Memory Recurrent Neural Network", 2015 IEEE International Conference on Smart City, pp. 153-158, 2015.
- [10] M. Ball and R. M. Slyke, "Backtracking algorithms for network reliability analysis", *Ann. Discrete Mathematics*, vol. 1, pp. 49-64, 1977.
- [11] R. R. Boorstyn and H. Frank, "Large-scale network topological optimization", *IEEE Trans. Communications*, vol. COM-25, no. 1, pp. 29-47, 1977.
- [12] ITU-T, "Architectural framework for machine learning in future networks including IMT-2020", ITU white paper. *Communications, Rec. ITU-T Y.3172*, pp. 14-19, 2019.

AUTHORS



Zhou Wei Gang obtained a master's degree in communication engineering in 2012. He worked for China Mobile Guizhou Company, mainly engaged in network analysis, data processing, IT development and artificial intelligence research. He has published 8 papers (in Chinese) including "Research on VoLTE MOS Value Slicing Evaluation Method", "Supporting 4G Business Development Based on the "User-Network-Business" Three-dimensional Method" (Chinese), and published "A Language Conversion Method Based on Artificial Intelligence". He has been involved in five patents including a method and device for early warning and judgment of single-point operation of transmission services. At present, he mainly studies the application of artificial intelligence in the field of the telecom industry.



Qianyin Rao obtained a bachelor's degree in the telecom industry engineering in 2005, and worked in the field of China Mobile transmission and IP bearer network production management and optimization. He published a patent "A method of MPLS service application based on VPN nesting technology" and "A degradation of IP network link quality" "Methods of Detection and Protection", won 1 group-level and 3 provincial-level scientific and technological innovation awards in the field of transmission and bearing. At present, his main research is the automatic detection of the same route of the transmission network and automatic service opening.



Lin Guo received a bachelor's degree in engineering in 2005 and worked in China Mobile in charge of core network maintenance, independent research and development, and AI design. He has published 7 national patents, "A Smart Remote Video

Interaction System Based on AR", "A New Method for Visual Fiber Cable Fault Location and Presentation", "A method for locating the root cause of communication network equipment in the network management system. Methods" etc. The application of 3D modeling technology in wireless is currently being studied.



Ze Zhong Feng holds a Bachelor of Engineering degree and is a senior engineer of China Mobile Guizhou company, engaged in AI research, time series prediction, and is committed to algorithm optimization of network topology. He is also committed to big data, machine learning and deep learning.



Lin XI, received a bachelor's degree in software engineering in 2018. Currently he is working in Guizhou Mobile, responsible for IT R&D and AI capability research. Currently the main research is structured data mining.



Qian Deng received a bachelor's degree in 2017 in information and computing science and a bachelor's degree in software engineering. He has worked in the complex IT R&D and AI capability research field of China Mobile. He has published the paper "The Realization of A Penetration Testing Assistant Tool", and published the patent "A Language Recognition, Method, System and Device". Currently, he mainly researches the direction of data mining through machine learning.

NOTE: The patent titles of the above-mentioned papers are all translated in Chinese and published in Mainland China.