Prequalification of VDSL2 copper customers for G.fast services with artificial intelligence technology

Pang-Chen Liu, Shun-Kai Yang, Lung-Chin Huang, Huai-En Tseng, Fei-Hua Kuo, Tai-Chueh Shih
Telecommunication Laboratories
Chunghwa Telecom
Taoyuan, Taiwan, R.O.C.
{bliu, conan, leopoldhuang, w180112, philip2, shihtc }@cht.com.tw

Abstract—***In recent years, due to customers have higher requirements for 4K/8K video and high-speed internet, telecom operators have begun to deploy FTTH network , but found that it is generally difficult to deploy fiber to the home, so G.fast technology has been favored by most telecom operators around the world and have begun to actively deploy. For the most widely deploy VDSL2 line with maximum rate that can only provide 100M internet service, a intelligent and accurate G.fast 300M high speed service prequalification technology, is a major research topic for telecom operators to promote 300M high-speed internet service. This paper proposes to use AI machine learning to estimate the G.fast line rate by using VDSL2 line attenuation to meet the real-site provision needs of telecommunications operators.***

Keywords-G.fast;VDSL2;Vectoring;Kera;MLP;Reinforcement Learning

I. INTRODUCTION

Copper wire is the most popular medium of access network in the world currently. However, due to the rapid increase in user bandwidth demand in recent years and the strong competition from cable TV operators, optical fiber has become the first choice for operators in network deployment. As operators gradually deploy fiber to customer's home, they find that Fiber-To-The-Home (FTTH) will generally encounter construction problems such as cable-wire congestion, timeliness, and user wishes. ITU-T G.fast [1~2] copper wire technology are comparable to the fiber bandwidth. It has the characteristics of reducing FTTH construction costs and quickly providing users with high-speed bandwidth services as a solution to transition to full fiber network. The international market research institution, Omdia predicts that the number of global G.fast connections will reach 29 million in 2021, accounting for 3% of the world's fixed network broadband market [3]. It is foreseeable that G.fast will be a main driving force of global telecommunication market in recent years.

The main difference between VDSL2 and G.fast technologies is the frequency band that increased from 30MHz to 106MHz. Real-site experience shows that VDSL2 can only provide 140Mbps of maximum bidirectional total rate, but G.fast can reach 1Gbps bidirectional total rate. However, the bottlenecks of high frequency bands are the problems of wire attenuation and interference susceptibility. According to the ITU-T G.fast standard, the connection rate is 200Mbps at 200 meters and 500 Mbps at 100 meters. If it provides bandwidth above 500Mbps, the distance will be shortened in 100 meters. However, after the G.fast lines interfere with each other, the rate will be reduced by about 200Mbps. Therefore, the self-crosstalk cancellation technology [4] (also called Vectoring) must be used to eliminate the interference between G.fast lines, and the rate could be maintained close to about 500 Mbps.

In generally, operators use the traditional "provision distance" method when copper wire is used to provide the internet access services. It is based on distance between the customer's home and telecommunication facilities to determine whether it is available for the customer to apply the service. For example, the provision distance of G.fast 300M service is 100 meters. If the customer distance is less than 100 meters, it can be provisioned, otherwise it cannot be provisioned.

Some issues inside copper wire are customer line conditions (such as customer lines have T connections), external noise interference (such as radio interference), and changes over time. Therefore, if the traditional provision distance method is used to determine whether G.fast 300M service is available, there will be a problem for provision. It is an important goal to identify potential customers and customers who cannot provide services. As a result, operators may be able to increase revenue and not waste unnecessary maintenance costs.

Therefore, this paper discusses and proposes to use AI machine learning to estimate the G.fast wire rate for satisfying the operator's real-site provision of 300M service requirements. The arrangement of this paper is as follows. Section 2 discusses the dependence analysis of copper cable quality and performance. Section 3 discusses machine learning models and applications. Section 4 describes the results of AI estimation accuracy in real-site cases. Section 5 is conclusion.

II. The interdependence analysis of the loop quality and performance for copper cable

Because self-far-end-crosstalk cancellation (Vectoring) technology was implemented, the self-FEXT would be dramatically decreased. The impact of the crosstalk from the neighborhood wire-pairs would be reduced for the loop performance, and the interdependence of the loop quality and performance would be more obvious. After artificial intelligence (AI) module learned the interdependence, the evaluation result of the field loop quality through the AI module analysis and classification could be utilized for the reference of the real-site provision.

A. Copper loop quality

Since loop attenuation gradually deviates from the reference model of normal loop as length, aging, and line problems, the loop qualities of the wire-pairs would have more difference. VDSL2[5] uses "Hlog" as loop attenuation and longer loop length or bad loop quality has worse Hlog. VDSL2 defines 17a and 30a profiles using spectrum bandwidth of 17.66MHz and 30MHz per user service bandwidth, and the subcarriers as transmission unit are 4.3125kHz for 17a and 8.625kHz for 30a. 17a and 30a would have 4,096 and 3,478 subcarriers, respectively. All Hlog per subcarrier would be condensed as 512 groups for information transfer of management, and the used "group size" including 1, 2, 4, and 8 would be decided by the amount of the actual active subcarriers. The Hlog of the loop with bridge tap shown as Fig. 1 had some depressions at specific frequency, and the line problems like bridge tap would make the loop quality worse.

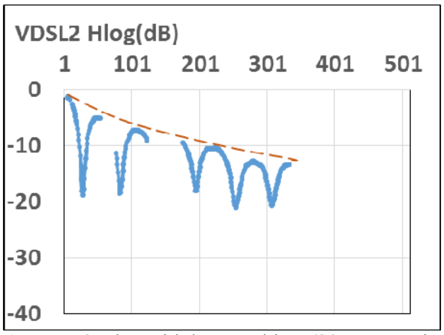

Fig. 1. VDSL2 Hlog with loop problem (30a, group size=8)

B. The interdependency analysis of loop quality and performance

According to the loop model Hlog(f)=αf^{β} (f is frequency), the Hlog of high frequency is related to the Hlog of low frequency except for specific line problems. It is expected to estimate the G.fast Hlog of 106MHz spectrum from the Hlog of VDSL2 17a or 30a, as shown in Fig. 2. The loop performance of G.fast is decided by the transmitted signal strength, loop attenuation (Hlog), crosstalk, and G.fast protocol parameters, and among them, loop attenuation is mainly external factor. The downstream GeoLATN[6] of VDSL2 is defined as the average of the Hlog for the downstream subcarriers, and could be used to analysis the interdependence with G.fast loop performance (aggregation net data rate(NDR) is the sum of NDR of downstream and upstream). When loop attenuation became worse and downstream GeoLATN of VDSL2 would be increased, the corresponding G.fast loop performance would decrease and its variation is about 100 to 150Mbps, as shown as Fig. 3. AI module learns the working model of Hlog estimation of G.fast from Hlog of VDSL2, G.fast transmitted signal strength, and G.fast protocol parameters. In Fig. 4, Hlog of VDSL2 with increasing loop distance from top to bottom is corresponding to the decreasing G.fast aggregation NDR that is classified as label 9 to 0 from 750Mbps to 350Mbps (50Mbps per label) shown as different color. We had built a draft AI module to evaluate the VDSL2 line, for an example service rate of 300Mbps and 100Mbps of downstream and upstream, there are 95% to match the suggestions of AI module estimation.

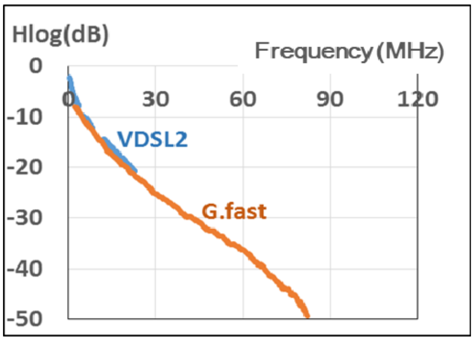

Fig. 2. Hlog of VDSL2 and G.fast

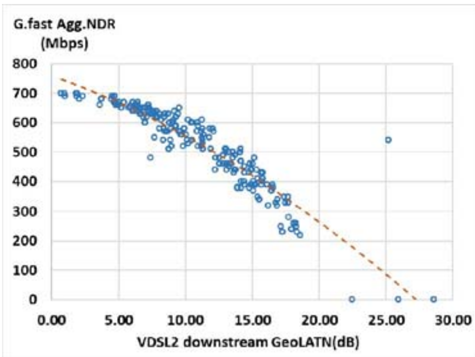

Fig. 3. Interdependence between downstream GeoLATN of VDSL2 and G.fast loop performance

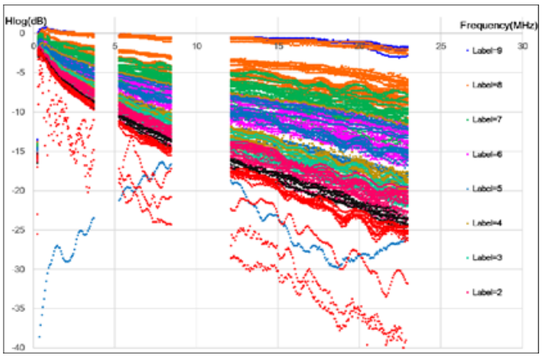

Fig. 4. Interdependence of VDSL2 Hlog and G.fast loop performance

III. ESTABLISHING PREDICTIVE MODELS AND APPLICATIONS

There are many tools for deep learning, for instance, TensorFlow, Keras, CNTK, and Theano. For model applications, our research uses TensorFlow and Keras, which are powerful, high efficiency and support various platforms. During study, we tried to use CNN and MLP multiple models to verify the Hlog data and various platform models. The Keras platform and MLP model are the most suitable for this research combination, both the prediction accuracy and loss rate are close to the actual G.fast line provision capacity. The following descriptions of the practical applications are all written by the combination of the Keras platform and the MLP model. Fig. 5 shows the entire AI model process architecture, which is explained in the following sections.

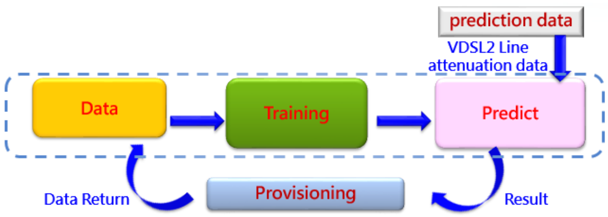

Fig. 5. AI model process

A. Data preparation

For the frequency field of Hlog , we split all data of Hlog into 512 intervals, and each interval is distributed from 0 to 96. Therefore, each line has 512 feature values, and each feature value is 0 to 96 value distribution. We define each G.fast rate to be divided into 10 equal label values, as shown in Table 1. The standardization of feature value is between 0 and 1, and the original data is converted into dimensionless scalar quantity [7]. Standardized data can improve the convergence speed of the model and the accuracy of the model. The label field is originally a number from 0 to 9 (such as the label value in Table 1), and shall be converted into 10 bits that is 0 or 1 combinations through one-hot encoding. For example, the tag value of 7 is 0000001000 after one-hot encoding conversion, corresponding to 10 neurons in the output layer.

We have collected 20,000 raw data, 80% of which are used to train the model, and 20% are used to validation dataset, which is used to calculate the accuracy of the trained model. According to the accuracy of our validation dataset and the actual prediction results, 80% training dataset and 20% validation dataset would be the optimized combination.

Table 1. Definition of feature values

Label Values	Downstream rate	Upstream rate	Whether 300M service is available
0	0	0	Unprovision
1	206	94	
2	240	110	
3	274	126	
4	309	141	
5	343	157	Provision
6	377	173	
7	411	189	
8	446	204	
9	480	220	

B. Modeling, training and prediction

We establish a multi-layer perceptron (MLP) neural model, and design two hidden layers (h1 and h2) in this system. It trains 80% of the dataset, so there are 16,000 training dataset. After the final training, the accuracy rate is approximately 0.94 and the error rate is approximately 0.11, as shown in Fig. 6.

A single-layer neural network can only be used to represent linearly separable functions, and it is explained in Section 3.2 that the data is non-linear and the data is in discrete form. After research, two hidden layers are used to establish the MLP model. The hidden layer with two activation functions, such as logistic sigmoid, in MLP that is enough to create a classification area of any desired shape, the MLP with a hidden layer is a universal approximation. Therefore, by using these two hidden layers for long-term verification test, we obtained the best accuracy for the entire prediction.

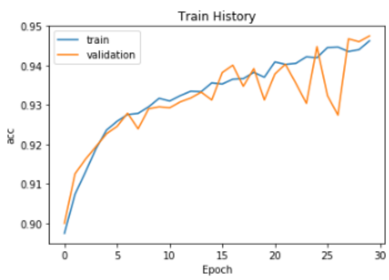

Fig. 6. Accuracy of the training process

C. Reinforcement learning

The new Hlog data which can be collected from different VDSL2 line is used for model prediction, according to the definition of the label value in Table 1. The prediction result is shown in Fig. 7. The probability field is the prediction result which is the number of label 0 ~ 9. As the features defined in Table 1, labels above 5 are available for provision, so it is displayed in the Option field whether they are available for provision.

In the preparation stage of AI neural data, the accuracy of definition features in training dataset will affect the accuracy of the prediction results. Therefore, in order to allow the AI model to reinforce and improve accuracy, when the prediction results are actually provided for G.fast 300M service, the results after prediction will be put into the AI model to train the model again. And the reinforcement learning[8] data is for actual provision, and the accuracy of AI will be close to the status of actual provision. After practical observation, the actual results after provision are reinforcement learning to the AI model, the accuracy has been slightly improved.

	File	probability	Down Stream	UP Stream	Option
0	DS-hlog_17a_sample0001.txt	8	446	204	Enable
1	DS-hlog_17a_sample0004.txt	8	446	204	Enable
2	DS-hlog_17a_sample0023.txt	6	377	173	Enable
3	DS-hlog_17a_sample0031.txt	5	343	157	Enable
4	DS-hlog_17a_sample0037.txt	0	0	0	Disable
5	DS-hlog_17a_sample0041.txt	4	309	141	Disable
6	DS-hlog_17a_sample0045.txt	4	309	141	Disable
7	DS-hlog_17a_sample0051.txt	2	240	110	Disable
8	DS-hlog_17a_sample0121.txt	2	240	110	Disable
9	DS-hlog_17a_sample0131.txt	9	480	220	Enable

Fig. 7. Forecast results data

IV. PREDICTION VERIFICATION

After training the model, we use twenty thousand raw data to predict. The accuracy of result is extremely high as 95%. It is same as our training model. We use confusion matrix[9] to reveal the result of prediction in Fig. 8.

Predict	0	1	2	3	4	5	6	7	8	9
Label										
0	1200	0	0	0	0	0	0	0	0	0
1	0	1021	0	0	0	0	0	0	0	0
2	0	80	1540	20	0	0	0	0	0	0
3	0	0	60	2101	0	0	0	0	0	0
4	0	0	0	80	1201	180	0	0	0	0
5	0	0	0	0	179	2211	0	40	0	0
6	0	0	0	0	0	40	2481	30	0	0
7	0	0	0	0	0	0	120	3180	30	0
8	0	0	0	0	0	0	0	40	3470	10
9	0	0	0	0	0	0	0	0	0	691

Fig. 8. The confusion matrix after prediction

To verify 300M high speed internet whether is work in real site, we chose 200 VDSL2 lines in real site. First, we input Hlog of the 200 lines to predict and accuracy of AI model is about 94%. Then we change these 200 VDSL2 lines to G.fast and verify whether internet speed can reach 300M and compare the statistic result. We find that the precision rate is 96%. The result can prove that AI prediction model accuracy is close to the precision rate in real site and after changing VDSL2 to G.fast to achieve 300M high speed internet is work. Table 2 shows the statistic of prediction and actual provision.

Table 2. The statistic of prediction and real environment provision

	Verification amount	Precision rate of AI model	Statistic of prediction and real environment provision	Error rate of AI prediction
Real site	200	94%	96%	2%

V. CONCLUSIONS

FTTH is actually the best way to satisfy strong demand of high-speed internet, but it usually faces to some difficult problems. The ITU-T G.fast, the rival of fiber channel, can decrease the construction cost of FTTH, provide high speed internet and also be a fast supplementary method to evolve to full-fiber network. In this paper, we propose to predict network speed in G.fast by estimating the loop attenuation of high frequency from legacy VDSL2 network and using machine learning. Operator can find more potential subscribers to subscribe 300M high speed internet in legacy VDSL2 subscribers and make high speed broadband service can be used in those environment which hardly install fiber.

In real site, we combine the AI model with internal operation system to make users in our company use the system to predict provision result conveniently. To decrease development cost and security issue, we migrate the AI model to the operation system. The operation system uses operating system command to call AI model to produce prediction result and information and also ask AI model training again from the result. Therefore, the AI model can train consistently and make precision rate increase.

REFERENCES

[1] ITU-T Recommendation G.9700," Fast access to subscriber terminals (FAST) - Power spectral density specification", April, 2014.

[2] ITU-T Recommendation G.9701," Fast Access to Subscriber Terminals (FAST) – Physical layer specification ", December, 2014.

[3] https://kknews.cc/zh-tw/tech/8842ln.html.

[4] ITU-T Recommendation G.993.5," Self-FEXT cancellation (vectoring) for use with VDSL2 transceivers ", February, 2019.

[5] ITU-T Recommendation G.993.2," Very high speed digital subscriber line transceivers 2 (VDSL2) ", February, 2019.

[6] Driton Statovci, "Prequalification of VDSL2 Customers for G.fast Services", 2017 25th European Signal Processing Conference (EUSIPCO)

[7] Chen Z Y, Ma N Z, Liu B. Lifelong learning for sentiment classification. In:Proceedings of ACL Conference. 2015

[8] Caruana R. Multitask learning. Machine Learning, 1997, 28(1)

[9] Pan S J, Yang Q. A survey on transfer learning. IEEE Transaction on Knowledge and Data Engineering, 2010, 22(10): 1345–1359