IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 24, 2019, accepted March 17, 2019, date of publication March 25, 2019, date of current version April 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2907142

A Supervised Learning Based QoS Assurance Architecture for 5G Networks

GUOSHENG ZHU[1], JUN ZAN[2], YANG YANG[1], AND XIAOYUN QI[1]
[1]School of Computer and Information Engineering, Hubei University, Wuhan 430062, China
[2]Hubei Provincial Safety Production Emergency Rescue Center, Wuhan 430064, China

Corresponding author: Xiaoyun Qi (346352@qq.com)

This work was supported in part by the Hubei Province Technology Innovation Project under Grant 2018ADC049.

ABSTRACT The 5G networks are broadly characterized by three unique features: ubiquitous connectivity, extremely low latency, and extraordinary high-speed data transfer. The challenge of 5G is to assure the network performance and different quality of service (QoS) requirements of different services, such as machine type communication (MTC), enhanced mobile broad band (eMBB), and ultra-reliable low latency communications (URLLC) over 5G networks. Unlike the previous ''one size fits all'' system, the softwarization, slicing and network capability exposure of 5G provide dynamic programming capabilities for QoS assurance. With the increasing complexity and dynamics of the network behaviors, it is non-trivial for a programmer to develop traditional software codes to schedule the network resources based on expert knowledge, especially when there is no quantitative relationship among the network events and the QoS anomalies. Machine learning is a computer technology that gives computer systems the ability to learn with data and improve performance and accuracy of decision making on a specific task, without being explicitly programmed. The areas of machine learning and communication technology are converging. Supervised learning based QoS assurance architecture for 5G networks was proposed in this paper. The supervised machine learning mechanisms can intelligently learn the network environment and react to dynamic situations. They can learn from the fore passed QoS related information and anomalies, and further reconstruct the relationship between the fore passed data and the current QoS related anomalies automatically and accurately. They, then, can trigger automatic mitigation or provide suggestions. The supervised machine learning mechanisms can also predict future QoS related anomalies with high confidence. In this paper, a case study for QoS anomaly root cause tracking based on decision tree was given to validate the proposed framework architecture.

INDEX TERMS 5G, architecture, quality of service, supervised learning.

I. INTRODUCTION

The 5G networks are expected to provide optimized support for a variety of services, such as different traffic loads, and end user communities [1]. The key performance indicators (KPIs) of 5G include high data rates, high user density, high user mobility, highly variable data rates, and coverage. The challenge of 5G is to assure the network performance and different Quality of Service (QoS) requirements of different services such as Machine Type Communication (MTC), enhanced Mobile Broad Band (eMBB), Ultra-Reliable Low Latency Communications (URLLC), Virtual Reality (VR)/ Augmented Reality (AR), 4K video streaming, multi-view 3D live streaming, Vehicle-to-Everything (V2X) [2], [3] and Machine to Machine (M2M) over the 5G networks. To meet the requirements of complex QoS of different applications and services, 5G networks are required to support the following QoS capabilities:

- QoS related events happened in the past can be reconstructed automatically and accurately;
- Current QoS related events can be detected accurately to trigger automatic mitigation or suggested immediate actions;
- Future QoS related events can be predicted with high confidence.

Unlike the previous 'one size fits all' system, the softwarization, slicing and network capability exposure of 5G provide dynamic programming capabilities for QoS assurance. With the increasing complexity and dynamics of network

The associate editor coordinating the review of this manuscript and approving it for publication was Huan Zhou.

behaviors, it is extremely hard for programmer to develop traditional software codes to schedule the network resources based on expert knowledge, especially when there is no mathematically causal relationship among the network events and the QoS anomalies.

The machine learning method, which can automatically drives algorithm optimization through data, has achieved great success in the fields of image/speech recognition, and automatic driving [4]–[6]. The application of machine learning in the network field has also become a research hotspot [7]–[13]. The design and management of networks and communication components can be significantly enhanced when combined with advanced machine learning methods. In particular, communication networks nowadays generate a huge amount of data at the network infrastructure level and at the user/customer level, which contain a huge amount of useful information such as location information, mobility and call patterns [14]–[21]. To improve the network performance and enhance users' experiences, new machine learning methods for big data analytics in communication networks can extract relevant information from the network data while taking into account of limited communication resources, and then leverage the derived knowledge for autonomic network control and management as well as service provisioning [22]–[24]. Machine learning is also proved to be applicable for automatic network orchestration and network management [25].

In this paper, a supervised learning based QoS assurance architecture for 5G networks was proposed. The supervised machine learning mechanisms, with the capabilities of teaching a computer to learn knowledge and concepts using data without being explicitly programmed, can intelligently learn the network environment and react to dynamic situations. They can learn from the past QoS related information and anomalies and reconstruct a relationship automatically and accurately. After the completion of training process that based on the past data, they can detect the current QoS related anomalies. They, then, can trigger automatic mitigation or suggest actions. Machine learning mechanisms can also predict future QoS related anomalies with high confidence.

This paper is organized as follows. In section II, we described the capabilities and requirements of 5G which include: network slicing, efficiency, multiple access technologies, QoS and policy control and performance requirements. Then we gave an introduction of machine learning which will be used in communication field in section III. The related work and challenges are given in section IV and V. The supervised learning based QoS assurance architecture was proposed in section VI. We finally present cases study of decision tree based 5G QoS anomaly detection and root cause tracking to validate the proposed framework architecture.

II. CAPABILITIES AND REQUIREMENTS OF 5G

A clear driver for 5G was the enormous growth of mobile data. It is reported that between 2020 and 2025 at the start of 5G mobile data traffic will have grown 1000x from 2010, at the start of 4G [26]. 5G technology will have to be able to support network operators to cope with that data growth. It is clear that a 1000 times growth of data volume cannot lead to a similar growth of costs of energy consumption; 5G will have to be more efficient than earlier generations.

The new capabilities of 5G for supporting ubiquitous connectivity, extremely low latency, and very high-speed data transfer are as follows:

A. NETWORK SLICING

A network slice is a logical network that provides specific network capabilities and network characteristics. Network slices enable the creation of customized networks to provide flexible solutions for different market scenarios which have diverse requirements, with respect to functionalities, performance and resource allocation. A network slice can provide the functionality of a complete network, including radio access network and core network functions. With the concept of network slicing, operators can customize their network for different applications and customers. One network can support one or several network slices.

B. EFFICIENCY

5G needs to be optimized for different service requirements. For Internet of Things based applications, optimizations are needed to handle very large numbers of devices. Configuration, deployment, and use of IoT devices may benefit from optimizations such as bulk provisioning, resource efficient access, and optimization for device originated data transfer. 5G will also have to efficiently support short data bursts without the need for lengthy signaling procedures before and after sending a small amount of data. 5G will also have to minimize signaling overhead and optimize access for user equipment with different mobility management needs. Applications such as voice telephony rely on the network to ensure seamless mobility. Applications such as video streaming on the other hand have application layer functionality (e.g. buffering) to handle service delivery interruptions during mobility. These applications will still require the network to minimize the interruption time.

C. COMMON CORE NETWORK WITH MULTIPLE ACCESS TECHNOLOGIES

The 5G network architecture is envisioned to be access network-agnostic, and with a core network common to radio access technologies (RATs), as well as existing fixed and wireless networks. The 5G core network should be accompanied by common control mechanisms that are decoupled from the access network technologies. The 5G system will support multiple 3GPP access technologies. Next to one or more 5G New Radio (NR) variants, it will also include 4G radio interface technology (E-UTRA). Furthermore, 5G will support various non-3GPP access technologies. 5G shall also be able to provide services via satellite access and fixed broadband access.

D. E2E QOS AND POLICY CONTROL

As 5G is expected to operate in a heterogeneous environment with multiple access technologies, multiple types of devices, etc., it should support a harmonized QoS and policy framework that applies to multiple accesses. Furthermore, 5G QoS needs to be end-to-end (including radio access, backhaul, core network, and network to network interconnect) to achieve the 5G user experience (e.g., ultra-low latency).

E. PERFORMANCE REQUIREMENTS

5G performance requirements highly depend on traffic scenarios. In an indoor hotspot scenario (e.g. for an office), the focus is on providing high data rates and high capacity. In a rural scenario, the focus is more on providing coverage.

The data rates for a rural scenario will be lower, but 5G should ensure that a minimum data rate is available also in urban and rural macro scenarios. Network access also needs to be supported in more extreme scenarios, with long range coverage, or in low end market scenarios, where access to power and backhaul facilities are not a given. Very large cell coverage areas of more than 100 km radius shall be supported with 1 Mbps downlink at cell edge. For constrained circumstances, and even larger areas, 5G shall be able to support a minimum user experience with 100 kbps, end-to-end latency of 50 ms, and a lower availability of 95%.

For vertical applications, other performance requirements are more important than data rates. For industry applications, the end-to-end latency is crucial. Motion control will not work if the time it takes to send information from a sensor to a controller is too long. Reliability – the percentage of packets successfully delivered within the time constraint – and communication service availability – the percentage of time the end-to-end communication service is delivered according to an agreed QoS – are crucial requirements for many industrial applications. Low latency requirements come with specific service area dimensions as low latency communication is only possible when source and destination are nearby. The speed of light in optical communication – approximately 1ms per 200 km – makes that network transport is not negligible, certainly not when latency caused by routers, switches and servers is taken into account. For the very low latency requirements of motion control, the communication from a sensor in the factory to a controller should be located within the same factory.

III. MACHINE LEARNING

Machine learning is a computer technology that gives computer systems the ability to learn with data and improve performance and accuracy of decision making on a specific task, without being explicitly programmed.

For a long time, the networking and distributed computing system is the key infrastructure to provide efficient computational resources for machine learning. Networking itself can also benefit from this promising technology [27], [28].

A. MACHINE LEARNING CLASSIFICATION

The goal of machine learning is to model an unknown target concept from observations. Depending on the nature of these observations, three types of learning can be identified [29]–[31]:

- The supervised learning: the observations are given in the form of input-output pairs. The purpose is to learn a function explaining the relationship between the inputs and outputs. Machine learning algorithm makes predictions on given set of sample whereas supervised learning algorithms searches for patterns within the value labels assigned to data points. This algorithm consists of an outcome variable which is to be predicted from, a given set of predictor's i.e. independent variables. Using these set of variables, we generate a function that map input to desired outputs. The training process continues until the model achieves level of accuracy on the testing data. Examples of supervised learning include: Neural Networks, Regression, Decision tree, k-Nearest Neighbor, Logistic Regression, Support Vector Machine, and Naive Bayes etc.
- The unsupervised learning: the observations were presented only the input values. The purpose is to find similarities between these values and to group similar data into clusters. In this algorithm, we don't have any target variable to estimate means and we don't have any label associated with data points or we can say class label of training data are unknown. This algorithms is used for organizing the data into the group of clusters to describe its structure i.e. cluster the data to reveal meaningful partitions and hierarchies. It makes data look simple and organized for analysis. Examples of unsupervised learning include: K-means, Fuzzy clustering, Hierarchical clustering.
- The semi-supervised learning where the observations are in the form of input-output pairs but the outputs values are not known in a large amount of observations. The purpose is to use these latter to improve the concept modelling in supervised learning.

Among these three types of learning, supervised and even semi-supervised learning best fit the QoS assurance modelling for 5G networks.

B. MODEL TRAINING

The model training can be offline or online. In offline learning, the model is first trained on a set of collected data, then, deployed on the new data. However, the online learning doesn't separate the model training and its deployment. It continues to learn as soon as new data are available in order to improve its performance. Thus, the online model can be adjusted very precisely to the end user.

The model training can be done in two ways: batch or incremental. In batch learning, data are predefined and available a priori. However, in incremental learning, data are introduced progressively and the model is able to change and adjust its settings after observation of each data in order to enrich their previously acquired learning from old data. The 5G QoS assurance models are generally designed to be used in real time and in dynamic environments. Thus, an incremental learning seems to be a promising solution. Indeed, it can easily be made online, enabling to use the model during it learning and to improve it throughout its use.

IV. RELATED WORK

QoS assurance is the degree of confidence that the process or deliverable meets defined characteristics or objectives. Machine learning is based on algorithms, data and computing power. Communication networks with distributed architecture and limited computing power have not been designed to cope with big data analytics and machine learning.

The machine learning algorithms which are applied to 5G QoS assurance could be distributed, centralized or hybrid which are bound by constraints on communication resources and operate in a noisy and dynamic environment. 5G networks are required to support diverse performance requirements of diverse services based on unified and E2E QoS control mechanisms. The data collected for 5G QoS assurance machine learning algorithms should include the User Equipment (UE) data, Access Network (AN) data, Core Network (CN) data and QoS KPI monitoring data.

A framework of big data driven mobile network optimization was proposed in [14]. The big data were collected not only from user equipments but also from mobile networks. Moreover, several techniques in data collection and analytics were discussed from the viewpoint of network optimization. Certain user cases on the application of the proposed framework for improving the network performance were also given in order to demonstrate the feasibility of the framework.

A machine learning engine was used to establish the relationship between the influencing factors and the QoE through artificial intelligence. For some typical approaches such as the neutral network model used in the Pseudo-Subjective Quality Assessment (PSQA) assessment method the model has to be trained before being used for performance evaluation. The analysis of large data sets leads to insights into the users' real experience, which may need to incorporate social data. With the integration of the 5G emerging mobile networks with Big Data analytics, the quality of mobile life was expected to be tremendously enhanced.

Another data-driven and machine learning architecture for enhancing personalized QoS/QoE for 5G networks was proposed in [15]. They specifically proposed a two-step QoE modeling approach to capture the strength of the relationship between users and services. Thereafter, the preferences of a user was introduced to model the user's subjectivity toward a specific service. With the comprehension of users' preferences, radio resources could be distributed more precisely. Simulation results showed that overall QoE can be enhanced by 20 percent, while 96 percent of users have an improved QoE, which validates the efficiency of the proposed architecture.

The complexity and dynamic of 5G network will produced massive and heterogeneous data, the unified data format is necessary for the data driven and machine learning model to work efficiently.

The 5G network management and orchestration plane need to detect the QoS/performance fault and anomalous events, then perform the management process aiming at fulfilment, assurance, and billing of services, network functions, and resources in both physical and virtual infrastructure including compute, storage, and network resources, or perform the orchestration aiming at the automated arrangement, coordination, instantiation and use of network functions and resources by optimization criteria [32].

V. CHALLENGES OF MACHINE LEARNING BASED QOS ASSURANCE FOR 5G NETWORKS

A. CHALLENGES FROM 5G USERS

A consistent user experience is expected in all types of scenarios, including in office towers, dense residential areas, stadiums, subways, highways, and so on. These scenarios, which are characterized by ultra-high traffic volume density, ultra-high connection density, or ultra-high mobility, are challenging for QoS management in 5G. For example, in high speed scenarios, maintaining a satisfactory level of service continuity is of great importance. Similarly, in the ultra-dense small cells scenario, designing an appropriate mobility management strategy in order to reduce the number of handovers, is a key issue to improve QoS. Therefore, maintaining a consistent and optimum user experience in such complicated network scenarios requires close coordination of the various QoS/QoE control strategies listed above.

B. CHALLENGES FROM 5G APPLICATIONS

The applications in 5G will much richer than those in 2G, 3G and 4G. With development in 5G, applications such as virtual reality, 3D videos, and interactive games may emerge. These emerging applications could bring challenges in the field of QoS/QoE. Since most existing models focus on VoIP, video, and WEB services, the new service characters and user demands should be studied in order to set up a proper QoS/QoE model. Imagine a medical monitoring service running on wearable equipment, which monitors blood pressure and pulse and even identifies when the user has an irregular heartbeat. In this case, the model should combine the individual user's information to make accurate decisions to achieve a high level of QoE.

C. CHALLENGES FROM 5G NETWORKS

In the 5G era, with the exponential growth in network data traffic and richer applications, it can be easily predicted that huge amounts of data will be generated in mobile communication networks. With the increasing complexity and dynamics of network behaviors, it is very hard to schedule the network resources based on expert knowledge, especially when there is no mathematically causal relationship among

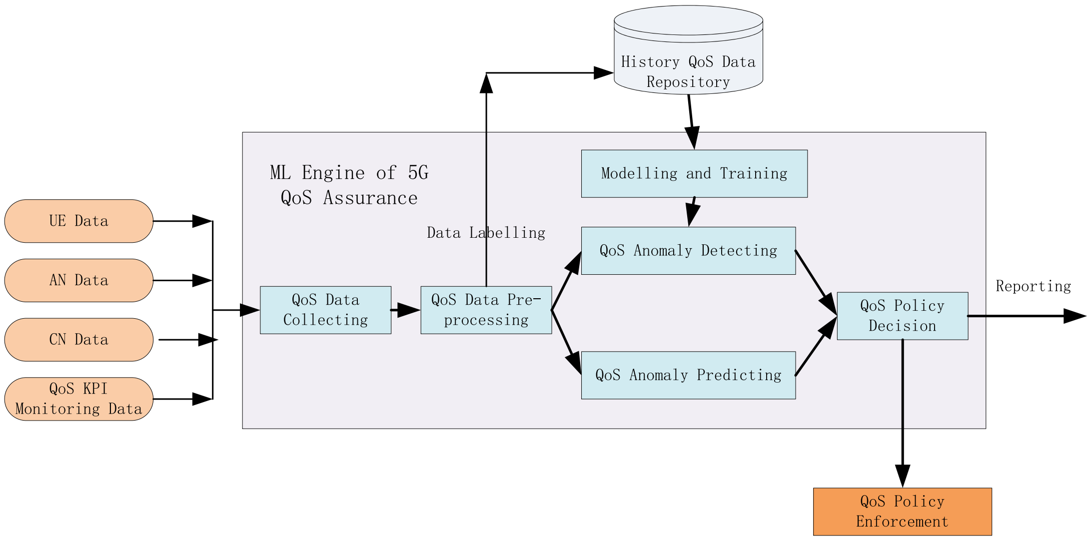

FIGURE 1. Architecture of supervised learning based QoS assurance for 5G networks.

the network data and the QoS anomalies. It is also challenging to handle the security and privacy related to massive network data with personal information, contacts, download history, application usage records, system logs, etc.

VI. ARCHITECTURE OF SUPERVISED LEARNING BASED QOS ASSURANCE FOR 5G NETWORKS

An architecture of supervised learning base QoS assurance for 5G is proposed and illustrated in figure 1. The machine learning engine of 5G QoS assurance is a basic function entity for QoS data collecting, pre-processing, modelling and training, QoS anomaly detecting, QoS anomaly predicting and reporting.

The 5G QoS related data from UE, AN, CN and QoS KPI monitoring data are collected by the 5G QoS data collector. After QoS data pre-processing, native 5G QoS data are cleaned and transformed into unified format. Anomaly is a pattern in the data that does not conform to the expected behavior. In order to detect and predict QoS anomaly, we need the labelled 5G history QoS data and machine learning based modelling/training. The modelling and training outputs include the QoS anomaly detector and predictor. The current and unified 5G QoS data are input to the 5G QoS anomaly detector and predictor, 5G QoS decision can be made by the QoS decision function entity and report the QoS results to 5G controlling and management function entities.

A. 5G QOS DATA COLLECTION

The volume, velocity and variety of the data from both 5G UEs and communication networks will explode exponentially. This provides opportunities for Mobile Network Operators (MNOs) to understand the behavior and requirements of users, which in turn allows for intelligence real-time QoS assurance decision making in a wide range of applications.

The collection of QoS assurance data can be achieved from the UEs, the ANs and the CN. The 5G QoS data can be collected in passive or active method. The passive method uses non-intrusive tools which will not affect the supervised elements. The active method uses intrusive tools which will affect the supervised elements, e.g. adding field of timestamps, flags to packet header for to monitored packets.

- *UE QoS data*: the UE hardware and software, physical resources and virtual resources, installed applications, etc. dynamic information data of 5G UEs which include the UE position, moving direction, speed, running applications, usage of physical/virtual resources(CPU, memory, energy, etc.), signaling, log and alarm information of fault, configuration, accounting, performance and security etc.
- *AN QoS data*: static configuration information data of 5G ANs which including the distribution of base stations, channel, antenna pattern, Committed Information Rate (CIR), frequency band, propagation type bandwidth, etc. dynamic information data of 5G ANs which including the usage of usage of physical/virtual resources (CPU, memory, energy, etc.), signaling on the air, log and alarm information of fault, configuration, accounting, performance and security etc.
- *CN QoS data*: static information of 5G CN which including the user Service Layer Agreement (SLA), physical/virtual network resources and available network slice resources. Active network slice instances and physical or virtual resource (CPU, memory, energy, etc.) allocated to the network slices, signaling, log and alarm information of fault, configuration, accounting, performance and security etc.

B. 5G QOS DATA PREPROCESSING

The collected 5G data are from UEs, ANs and CN with different time, spatial and format features. In order to store and analyze the massive QoS data, it is required to pre-process

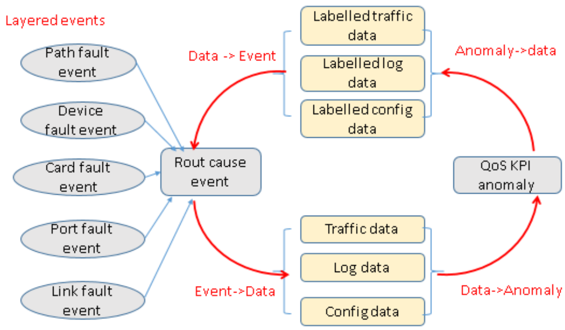

FIGURE 2. Event-data-QoS anomaly.

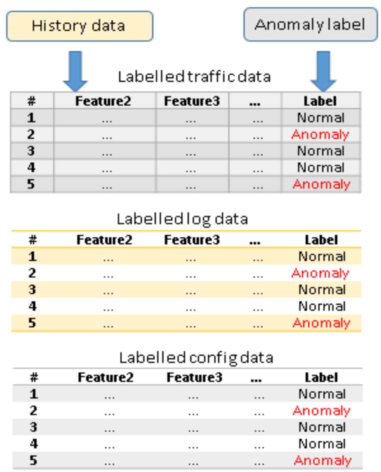

FIGURE 3. Collection and labelling of QoS related data.

collected data and transforms it into understandable and easy-to-use structures:

- Extract-transform-load (ETL) the collected multi-source, heterogeneous 5G QoS data and transforms it into understandable and easy-to-use structures.
- Clean and filter noisy data from the collected multi-source, heterogeneous 5G QoS data.
- Normalize and unify the data format of the collected multi-source, heterogeneous 5G QoS data for further repository and analysis.
- Data pre-processing can be implemented through distributed computing architecture with the capability of scaling.
- Label the QoS related anomalies and the corresponding collected 5G QoS information data when it is possible.

C. 5G QOS DATA REPOSITORY

Repository of massive 5G QoS data faces challenges of scalability, performance, volume, velocity, and variety.

The multi-source, heterogeneous 5G QoS data can be stored using database (DB) or file system (FS), e.g., relational database, NoSQL database, file system, distributed file

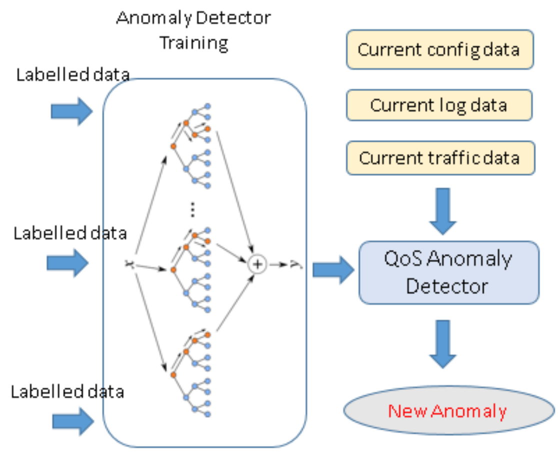

FIGURE 4. QoS anomaly detector and root cause tracking.

system. In our case study, we store 5G QoS related data in NoSQL database.

D. SUPERVISED MODELING AND TRAINING

Machine learning is based on algorithms, data and computing power. Based on the available 5G QoS data, the computing capabilities, we need to choose and construct appropriate model to correlate the collected 5G QoS data and QoS KPI parameters. The machine learning algorithms include supervised learning, un- supervised learning, semi-supervised learning, deep learning, reinforcement learning, etc. It is required to construct appropriate machine learning model(s) based on the collected 5G QoS information data and QoS KPI parameters.

We used supervised learning named C4.5 decision tree algorithm to construct model(s) between QoS data and QoS KPI parameters in our case study. The decision tree is a tree structure and can be a binary tree or a non-binary tree. Each of its non-leaf nodes represents a test on a feature attribute, each branch representing the output of the feature attribute over a range of values, and each leaf node storing a category. The decision process using the decision tree is to start from the root node, test the corresponding feature attributes in the item to be classified, and select the output branch according to its value until the leaf node is reached, and the category stored by the leaf node is used as the decision result.

E. QOS ANOMALY DETECTION AND PREDICTION

The machine learning based QoS modelling and training results include the QoS anomaly detector and predictor. The QoS anomaly detector can detect QoS anomaly in real time with high confidence which can be used for network/application/service trouble shooting and can trigger network resources re-scheduling. The QoS anomaly predictor can predict QoS anomaly which can trigger network optimization and planning. Another capability for QoS anomaly detection and prediction is the root cause tracking with many potential causes of QoS anomaly.

F. QOS POLICY DECISION AND REPORTING

The machine learning based QoS anomaly detection and prediction results can be used for resources re-scheduling, network optimization and planning. The results were illustrated in visual representations through various charts, tables and images, etc., so as to help MNOs to gain the meaning of big data apparently. The results can also be reported to other function entities in 5G network for resources re-scheduling, network optimization and planning.

G. QOS POLICY ENFORCEMENT

The 5G QoS enforcement functions are distributed in control plane and user plane. For example, the 5G network slice instance management and orchestration function, are in the control plane. The traffic/flow classification, marking, policing, queuing/scheduling and shaping functions in the user plane. The machine learning mechanisms can also be used in all these QoS policy enforcement functional entities.

VII. CASE STUDY OF ROOT CAUSE TRACKING OF QOS ANOMALY

A. LAYERED CAUSE OF QOS ANOMALY

The root cause tracking of network QoS anomalies is very important for services/applications QoS assurance. While the causes of QoS anomalies are layered, see the following list. The example network fault events include:

- Path fault event;
- Device fault event;
- Card fault event;
- Port fault event;
- Link fault event.

The phenomenon of network performance anomalies are network data, such as, network traffic data, syslog data and management data, etc. The QoS anomaly network data are from network anomaly events, such as, network attacks, protocol bugs and link up/down, etc. Based on the analysis of multi-layer dependence and spatial-temporal dependence of network data and network events, we can reversely track the root causes of network anomalies. It is possible to clarify the positive correlations and reverse tracking mechanisms of network 'anomaly events – anomaly data – network anomalies'.

B. LABELLING OF COLLECTED QOS DATA

The collected data are pre-processed and transformed into unified format with labels.

C. TRAINING OF QOS ANOMALY DETECTOR

The labelled data are stored in QoS data repository and used to train the C4.5 decision tree algorithm [15].

Given a set S of cases, C4.5 first grows an initial tree using the divide-and-conquer algorithm as follows:

1. If all the cases in S belong to the same class or S is small, the tree is a leaf labeled with the most frequent class in S.

2. Otherwise, choose a test based on a single attribute with two or more outcomes. Make this test the root of the tree with one branch for each outcome of the test, partition S into corresponding subsets S1, S2,... according to the outcome for each case, and apply the same procedure recursively to each subset.

There are usually many tests that could be chosen in this last step. C4.5 uses two heuristic criteria to rank possible tests: information gain, which minimizes the total entropy of the subsets {Si} (but is heavily biased towards tests with numerous outcomes), and the default gain ratio that divides information gain by the information provided by the test outcomes

The training result named QoS anomaly detector can be used to detect the current QoS anomaly when the online QoS data are input to the detector.

The QoS anomaly detector can detect the new anomaly with high confidence and the result can be stored in the QoS data repository to re-train and improve the model. The model is generally designed to be used in real time and in dynamic environments which makes this incremental learning be a promising solution.

To verify the accuracy of the model, we select five data sets with captured traffic data, see detailed information in Table 1.

We establishes a confusion matrix to verify the accuracy of anomaly detection architecture, see table 2.

TABLE 1. Sample data sets.

Data Set	Normal	Abnormal
Data set 1	1045973	3021
Data set 2	53514	503
Data set 3	1287134	3156
Data set 4	863695	2896
Data set 5	92119	735

TABLE 2. confusion matrix to verify the accuracy.

		Real Sample Value	
		Positive	Negative
Detection Value	Positive	True Positive(TP)	False Positive(FP)
	Negative	False Negative(FN)	True Negative(TN)

TABLE 3. Testing result of accuracy.

Data Set	Detected Anomaly	Accuracy (%)
Data set 1	1023	96.60
Data set 2	122	98.50
Data set 3	1074	96.62
Data set 4	1028	97.37
Data set 5	253	96.65

The detection accuracy can be computed using: $(TP + TN)/(TP + TN + FP + FN)$. Based on the five sample data sets, the testing results are illustrated in Table 3.

We can see the accuracy is more than 96%. Based on the anomaly detection and multi-layer dependence and spatial-temporal dependence of network data and network events, we can reversely track the root causes of network anomalies.

VIII. CONCLUSION

The challenge of 5G is to assure the network performance and different QoS requirements of different services such as Machine Type Communication, enhanced Mobile Broad Band, Ultra-Reliable Low Latency Communications over 5G networks. With the increasing complexity and dynamics of network behaviors, it is very hard for programmer to develop traditional software codes to schedule the network resources based on expert knowledge, especially when there is no mathematically causal relationship among the network events and the QoS anomalies. A supervised learning based QoS assurance architecture for 5G networks was proposed in this paper. The supervised machine learning mechanisms, with the capabilities of teaching a computer to learn knowledge and concepts using data, can intelligently learn the network environment and react to dynamic situations. They can learn past QoS related information and anomalies and reconstruct a relationship automatically and accurately. After completion of training based on the past data, they can detect current QoS related anomalies. They, then, can trigger automatic mitigation or suggest actions. The supervised machine learning mechanisms can also predict future QoS related anomalies with high confidence. Case study of decision tree based QoS assurance were given to validate the proposed framework architecture.

REFERENCES

[1] *Service Requirements for the 5G System, Stage 1,(Release 16)*, document TS22.261, 3GPP, 2018.

[2] H. Zhou, H. Wang, X. Chen, X. Li, and S. Xu, "Data offloading techniques through vehicular ad hoc networks: A survey," *IEEE Access*, vol. 6, pp. 65250–65259, 2018.

[3] C. Huang, Y. Chen, S. Xu, and H. Zhou, "The vehicular social network (VSN)-based sharing of downloaded Geo data using the credit-based clustering scheme," *IEEE Access*, vol. 6, pp. 58254–58271, 2018.

[4] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[6] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 6645–6649.

[7] C. Zhang, P. Patras, and H. Haddadi. (2018). "Deep learning in mobile and wireless networking: A survey." [Online]. Available: https://arxiv.org/abs/1803.04311

[8] Z. M. Fadlullah *et al.*, "State-of-the-art deep learning: Evolving machine intelligence toward tomorrow's intelligent network traffic control systems," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2432–2455, 4th Quart., 2017.

[9] Y. Sun *et al.*, "CS2P: Improving video bitrate selection and adaptation with data-driven throughput prediction," in *Proc. ACM SIGCOMM Conf.*, Aug. 2016, pp. 272–285.

[10] B. Mao *et al.*, "Routing or computing? The paradigm shift towards intelligent computer network packet transmission based on deep learning," *IEEE Trans. Comput.*, vol. 66, no. 11, pp. 1946–1960, Nov. 2017.

[11] J. Jiang, V. Sekar, H. Milner, D. Shepherd, and I. Z. H. Stoica, "CFA: A practical prediction system for video QoE optimization,"in *Proc. 13th USENIX Symp. Netw. Syst. Design Implement. (NSDI)*, 2016, pp. 137–150.

[12] J. Jiang, S. Sun, V. Sekar, and H. Zhang, "Pytheas: Enabling data-driven quality of experience optimization using group-based exploration-exploitation," in *Proc. NSDI*, vol. 1, Mar. 2017, pp. 393–406.

[13] Z. Chen, J. Wen, and Y. Geng, "Predicting future traffic using hidden Markov models," in *Proc. IEEE 24th Int. Conf. Netw. Protocols (ICNP)*, Nov. 2016, pp. 1–6.

[14] M. Ruan, X. Chen, and H. Zhou, "Centrality prediction based on k-order markov chain in mobile social networks," *Peer-to-Peer Netw. Appl.*, to be published.

[15] H. Zhang, Y. Qi, H. Zhou, J. Zhang, and J. Sun, "Testing and defending methods against DoS attack in state estimation," *Asian J. Control*, vol. 19, no. 3, pp. 1295–1305, Jul. 2017.

[16] K. Zheng, Z. Yang, K. Zhang, P. Chatzimisios, K. Yang, and W. Xiang, "Big data-driven optimization for mobile networks toward 5G," *IEEE Netw.*, vol. 30, no. 1, pp. 44–51, Jan./Feb. 2016.

[17] M. Musolesi, "Big mobile data mining: Good or evil?" *IEEE Internet Comput.*, vol. 18, no. 1, pp. 78–81, Jan./Feb. 2014.

[18] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[19] A. P. Muniyandi, R. Rajeswari, and R. Rajaram, "Network anomaly detection by cascading K-means clustering and C4. 5 decision tree algorithm," *Procedia Eng.*, vol. 30, pp. 174–182, Jan. 2012.

[20] R. K. Lomotey and R. Deters, "Towards knowledge discovery in big data," *Proc. IEEE 8th Int. Symp. Service Oriented Syst. Eng. (SOSE)*, Apr. 2014, pp. 181–191.

[21] C. Wang, H. Lin, and H. Jiang, "CANS: Towards congestion-adaptive and small stretch emergency navigation with wireless sensor networks," *IEEE Trans. Mobile Comput.*, vol. 15, no. 5, pp. 1077–1089, May 2016.

[22] L. Gu, D. Zeng, P. Li, and S. Guo, "Cost minimization for big data processing in geo-distributed data centers," *IEEE Trans. Emerg. Topics Comput.*, vol. 2, no. 3, pp. 314–323, Sep. 2014.

[23] L. Lei, Z. Zhong, K. Zheng, J. Chen, and H. Meng, "Challenges on wireless heterogeneous networks for mobile cloud computing," *IEEE Wireless Commun.*, vol. 20, no. 3, pp. 34–44, Jun. 2013.

[24] S. Samulevicius, T. B. Pedersen, and T. B. Sorensen, "MOST: Mobile broadband network optimization using planned spatio-temporal events," in *Proc. IEEE 81st Veh. Technol. Conf. (VTC Spring)*, May 2015, pp. 1–5.

[25] A. Ramaprasath, A. Srinivasan, and C.-H. Lung, "Performance optimization of big data in mobile networks," in *Proc. IEEE 28th Can. Conf. Elect. Comput. Eng. (CCECE)*, May 2015, pp. 1364–1368.

[26] Focus Group on Machine Learning for Future Networks Including 5G. (2018). *Term of Reference*. [Online]. Available: https://www.itu.int/en/ITU-T/focusgroups/ml5g/Documents/FG-ML5G_ToRs.docx

[27] T. Norp, "5G requirements and key performance indicators," *J. ICT* vol. 6.1, pp. 15–30, May 2018.

[28] M. Wang, Y. Cui, X. Wang, S. Xiao, and J. Jiang, "Machine learning for networking: Workflow, advances and opportunities," *IEEE Netw.*, vol. 32, no. 2, pp. 92–99, Mar./Apr. 2018.

[29] S. Khorsandroo and R. M. Noor, "A generic quantitative relationship to assess interdependency of QoE and QoS," *KSII Trans. Internet Inf. Syst.*, vol. 7, no. 2, pp. 327–346, Feb. 2013.

[30] S. Aroussi and A. Mellouk, "Survey on machine learning-based QoE-QoS correlation models," in *Proc. Int. Conf. Comput., Manage. Telecommun. (ComManTel)*, Apr. 2014, pp. 200–204.

[31] M. Elkotob, D. Grandlund, K. Andersson, and C. Ahlund, "Multimedia QoE optimized management using prediction and statistical learning," in *Proc. IEEE Local Comput. Netw. Conf.*, Denver, CO, USA, Oct. 2010, pp. 324–327.

[32] S. Aroussi, T. Bouabana-Tebibel, and A. Mellouk, "Empirical QoE/QoS correlation model based on multiple parameters for VoD flows," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Anaheim, CA, USA, Dec. 2012, pp. 1963–1968.

[33] *Requirements of Machine Learning Based QoS Assurance for IMT-2020 Network*, Standard Recommendation ITU-T Y.3170, 2018.

GUOSHENG ZHU received the B.S. degree in computer application, the M.S. and Ph.D. degrees in computer architecture from the Huazhong University of Science and Technology, China, in 1994, 2001, and 2010, respectively. He is currently a Professor with the School of computer and information engineering, Hubei University, China.

From 2001 to 2015, he was a Researcher with the Wuhan Research Institution of Posts and Telecommunications. His research interests include big data, machine learning, communication networks, and related standards.

His awards and honors include the National Technology Invention Award, in 2014, for "The LAPS method of converged transport of Ethernet and SDH," the Hubei Science and Technology Progress Award, in 2013, for "Key Technologies and Applications of Packet Multiservice Transport and Sensing."

JUN ZAN received the Ph.D. degree in safety technology and engineering from the China University of Geosciences, China, in 2010.

He has engaged in safety professional work for 20 years. As the person in charge, he developed the digitalization standardization system for safety hazard investigation, realized the dynamic management of 140 000 enterprise data and hidden danger rectification information, and presided over the construction of the provincial emergency rescue command center and platform.

YANG YANG received the B.E. and M.S. degrees from the Wuhan University of Technology, China, in 2009 and 2012, respectively, and the Ph.D. degree from the Huazhong University of Science and Technology, China, in 2017. He is currently an Assistant Professor with the School of Computer Science and Information Engineering, Hubei University. His research interests include wireless networks and mobile computing.

XIAOYUN QI received the B.S. and M.S. degrees in chemistry engineering from the Wuhan Institute of Technology, China, in 1997 and 2003, respectively, and the Ph.D. degree in mechanical engineering from Hubei University, China, in 2012.

Since 2003, she has been an Associate Professor with Hubei University. Her research interests include big data, machine learning, communication networks, and related standards.