

Research Article

A Data-Driven Customer Quality of Experience System for a Cellular Network

Hyunglok Jung,¹ Jeonghoon Mo,² and Jungju Park²

¹Graduate Course of Technology Policy, Yonsei University, Seoul, Republic of Korea

²Department of Industrial Engineering, Yonsei University, Seoul, Republic of Korea

Correspondence should be addressed to Jeonghoon Mo; j.mo@yonsei.ac.kr

Received 18 May 2017; Revised 8 August 2017; Accepted 27 August 2017; Published 26 October 2017

Academic Editor: L. J. García Villalba

Copyright © 2017 Hyunglok Jung et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Improving customer-perceived service quality is a critical mission of telecommunication service providers. Using 35 billion call records, we develop a call quality score model to predict customer complaint calls. The score model consists of two components: service quality score and connectivity score models. It also incorporates human psychological impacts such as the peak and end effects. We implement a large-sized data processing system that manages real-time service logs to generate quality scores at the customer level using big data processing technology and analysis techniques. The experimental results confirm the validity of the developed model in distinguishing probable complaint callers. With the adoption of the system, the first call resolution rate of the call center increased from 45% to 73%, and the field engineer dispatch rate from 46% to 25%.

1. Introduction

The desire to understand customer experience was stimulated by intense competition in the telecommunications industry. As the market saturated, thereby intensifying competition, service providers endeavored to increase customer retention. Customer churn, which can be led by customer dissatisfaction [1], is disadvantageous for service providers: it causes not only decreases in profit, but also the loss of opportunities for existing customers to suggest their service operator to other users. Moreover, attracting a new customer costs more than retaining an existing one [2]. For these reasons, improving customer experiences has been a critical factor in business.

Quality of experience (QoE) has been of intense interests both from academia and industry for the last decade because it can provide a more holistic view of customer experience than its technology-centric complement, quality of service (QoS) [3]. Especially, the implementation of QoE systems has been enabled by the development in computing technology such as big data analytics, cloud computing, and so on. Smart phones and devices have generated a large volume of data under an environment in which information can be stored at distributed storage, called cloud storage, at low cost. In addition, technology development in distributed data

processing, such as Apache Spark, enables data processing at low cost [4].

Telecommunication service providers have developed large-scale systems to estimate overall customer satisfaction or make efforts to improve it. In mobile communication customer surveys in Hong Kong, coverage and transmission quality were found to be important factors [5]. The relationships between customer satisfaction, switching cost, and customer loyalty in France were studied [6]. Customer satisfaction is discussed to be measured through sales, installation, product use, repair, and billing experiences [7, 8]. A model was developed to explain the overall perceived performance at the customer level based on billing, branch networks, fault repair, service provision, brand image, and product satisfaction [9]. Because these approaches rely on surveys, they can be used for a posteriori evaluation, but not for predictive purposes.

The telecommunication service providers manage customer relationships using various types of key performance indicators (KPIs) to evaluate the overall customer-perceived quality. Among the KPIs, customer dissatisfaction with a service is an important KPI for telecommunication service providers. There is a positive correlation between customer dissatisfaction and customer complaining behavior [10].

Experiencing dissatisfaction, customers may take several actions to dissuade their friends or acquaintances from using the service provider's products or services, or to complain through consumer advocacy groups [11].

More recent endeavors implement the QoS-/QoE-based system to measure the status of the network. A method using the operational data from user equipment and different network elements is proposed to overcome the challenges and limitations of manual drive testing [12]. Under the paradigm of big data, the probability of subscriber churn is supposed to be well predicted because the intersubscriber influence is adopted [13]. A network planning tool was developed using the estimated QoS scores based on user-collected measurements through supervised learning tools [14]. A SlopeOne model-based QoS prediction method is also proposed to overcome the difficulty of customer clustering of the popular collaborative filter algorithm [15]. For the proper management of mobile edge computing ecosystems, a reference architecture is proposed to leverage network functions virtualization and software-defined networking technologies [16].

Many of the studies use consumers' subjective quality experience to improve network operation and consumer relationship following the standard of ITU, mean opinion score (MOS, ITU-T P.800) [17] or the E-model (ITU-T G.107) [18]. MOS is a subjective quality evaluation model using the averages of individuals' subjective evaluations of the quality of telephone calls. The E-model was developed to calculate MOS using objective metrics such as delay time and packet loss rate for a session; this model is extended to consider video and data (ITU-T P.910) [19].

However, we address that these models may not show customers' overall experience scores; rather it evaluates only the session. It also fails to explain the differences between customers, as it is a prediction model based on logical inferences for consumer-perceived quality. In other words, this model may have a little predictive power regarding overall customer-perceived quality.

Among the various complaint actions, the complaint call is an index that can accurately show customer dissatisfaction. Yurtseven advocated using the complaint call as a proxy for the overall perceived quality [20]. The service level index (SLI) measures customers' perceived quality from individual experiences [21]. In this paper, we develop a model that is different from SLI with the objective of predicting the likelihood of complaint calls. Quick and effective responses to complaint calls can prevent customer churn [22].

To the best of our knowledge, little effort has been devoted to predicting the likelihood of complaint calls. This may be attributed to several difficulties in indexing the complaint call. First, many customers, who have experienced poor service quality, do not make complaint calls [10]. Moreover, complaints also depend on the type of goods where 49.6%, 29.4%, and 23.2% of customers take no actions for perishable goods, durable goods, and services, respectively. Second, not every service quality from the complaint calls is at a poor level. It means that customers sometimes make complaint calls despite they experienced a high QoS. For these reasons,

additional effort is required to develop a model to predict the likelihood of complaint calls.

Our contributions are summarized as follows:

- (i) We propose a customer score model using machine learning techniques where the target variable of customer complaint calls has significant relevance to customer-perceived quality. This method integrates a subjective quality model with an objective one; its quality evaluations for individuals are automated. The objective quality model consists of access and service quality models.
- (ii) We configure a real-time cross-call detailed record (cross-CDR) database using (telephone number, call ID) as a key, based on various CDRs from in-memory computing using the open source Spark software. This method should enable real-time customer quality evaluation, monitoring, and analysis for several billion customer level quality logs per day. We expect that this method will be used as a building block towards future self-organizing networks (SONs) (the autonomous SON requires feedback loop as an input. We think the customer score model could be a very valuable input to SON by signaling poor performing network blocks).

The remainder of the paper is organized as follows. In Section 2, we compare the KPI-based method, which is a popular quality management method, and the customer-experience-based method. In Section 3, we introduce the composition of the overall system. In Section 4, we introduce a method to develop the access quality score model, the service quality score model, and the subjective quality score model. In Section 5, we verify the proposed models and present the experimental results. Finally, we conclude the paper in Section 6.

2. CEI-Based Quality Management

Customer experience indicator (CEI) quality management is a new paradigm in telecommunication quality management to focus more on individual customers. Before the paradigm, the focus of quality management is more on network system and equipment.

2.1. Network Oriented Quality Management. To deliver excellent QoS, many telecommunication service providers adopt a quality management framework that consists of quality definition, data collection, analysis, and corrective actions [23]. They select and monitor various KPIs. KPIs can be at the service level or at the system equipment level, such as *success rate for radio resource control request* of evolved Node B (eNB), *paging success rate* and *attach success rate* of mobility management element (MME), and *bearer request success rate* of packet gateway (PGW). The thresholds for each KPIs are set, and various data from the network management system, protocol analyzers, and complaint calls are measured and collected. The collected data are compared against the thresholds and analyzed. If necessary, corrective actions are taken to improve quality of the network systems. KPI-based

quality control can be limited, because there can be too many KPIs to monitor as the size of the network increases. Moreover, many KPIs are correlated, which makes it difficult to determine the origin of any network problem may arise. In addition, KPIs do not necessarily reflect the user-level QoE, because they typically blur individuals' characteristics using averages for one part of the network.

2.2. Individual Oriented CEI-Based Quality Management.

The concept of QoE combines user perceptions, experiences, and expectations with nontechnical and technical parameters to better address customer needs [24]. Qualinet defines it as the degree of delight or annoyance of the user of an application or service [25]. For example, *customer experience management index* collects customers' experiences on the benefit, the convenience, and so on, from each unit of a network [26]. The CEI is different from KPIs in that it is an individual-level indicator that reflects different customer behaviors. The CEI concept attempts to incorporate customers' individual experiences, needs, and behaviors with the technical KPIs to promote the optimal use of available resources. CEI-based quality management can achieve high retention rates, produce favorable recommendations, and (potentially) stimulate more service use. This method can locate problems and make improvements to support the provision of acceptable service experiences to customers.

2.3. *Limitations of KPI-Based Systems.* The following issues can occur in current KPI-based quality management systems.

(i) *KPI Management Overhead.* Telecommunication service providers manage the KPIs of various networks such as wireless access, transport, and core networks. As the network becomes more complex, it may generate large overhead costs for network monitoring, troubleshooting, and improving KPIs. To ensure the best possible QoE on service, and to confirm that services are fit for purpose, it is necessary to manage many different KPIs in near-real-time. However, generally available systems and reports provide averages and do not cater to the experience of individual customers.

(ii) *Slow Response to Customer Complaints.* Managing KPIs at the device level or the system level may cause slow responses to individual customer complaint calls. The network engineers must determine the root cause of the problem after reception of a call without individual-level experience information. Additionally, the lack of quantitative data for customer-perceived quality can prevent the service provider from undertaking proactive actions. Misclassification of customer complaints may cause unnecessary dispatches to sites, which can be expensive. Approximately 46% of dispatches to sites are unnecessary (Figure 8 shows that, after the system adoption, the rate of the resolution by field engineer decreased from 46% to 25% ($1-0.25/0.46 = 0.46$)).

(iii) *Inefficient Use of Resources.* Without knowledge regarding customer experiences, the available resources might not be used where they are needed most. Understanding the individual perceived quality levels can be helpful for locating

problematic areas that require more investments or improvements. As budgets are limited, improvements can be applied in a suitable order.

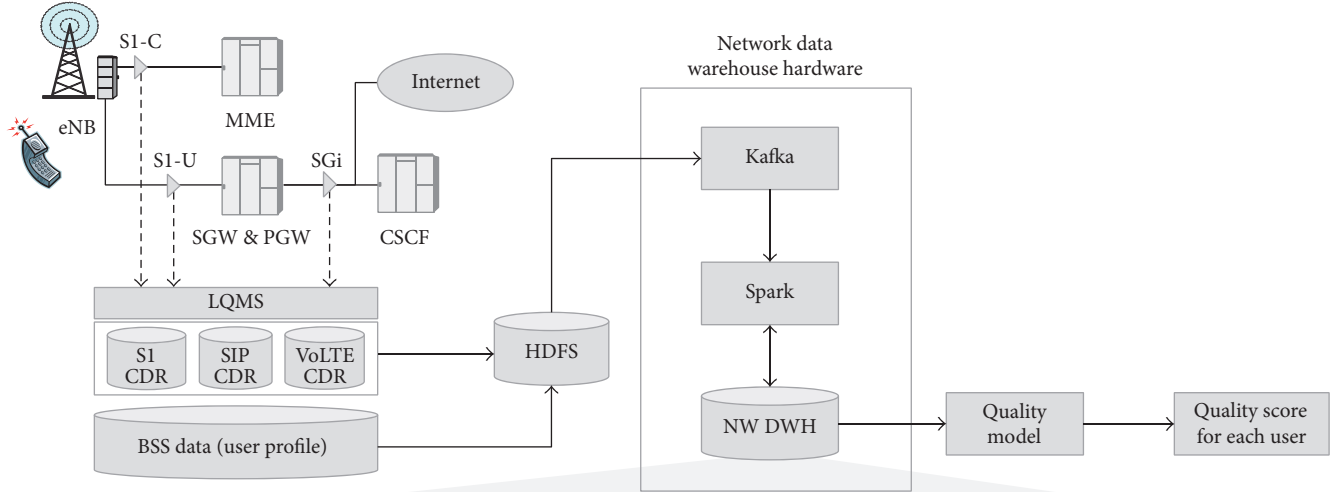
3. System Configuration and Data Collection

To implement a real-time individual-level perceived quality score system, we may need to assess tens of millions of customers making several billion calls per day. Figure 1 shows the system configuration that we implemented to generate a perceived quality score for each user.

The LTE network consists of eNB, MME, serving gateway (SGW), PGW, and call session control function (CSCF) with many interfaces. The eNB is the hardware that communicates directly with mobile phones. The MME, the key control-node for the LTE access network, is responsible for the authentication, service activation/deactivation, and so on. The SGW routes and forwards data packets and is involved in handover process as the mobility anchor. The PGW is the point of exit and entry to external data network. The CSCF deals with signaling of phone calls. For more details on the LTE system, refer to [27].

Three interfaces are shown in Figure 1. S1-C is the interface between eNB and MME for the control plane protocol. S1-U is the interface between eNB and SGW for the per bearer user plain tunneling and inter-eNB path switching during handover. SGI is the interface among PGW, CSCF, and the Internet. We implemented multiple tappers (three in Figure 1) for packet probing on the control plane (S1-C), the data plane (S1-U), and the SGI interfaces. The protocol analyzer decodes tapped packet streams and generates various CDRs. A single phone call generates multiple CDRs, S1 CDR, session initiation protocol (SIP) CDR, and VoLTE CDR, each of which is stored in different databases in the Hadoop distributed file system (HDFS). Every minute, newly generated CDRs are stored in the HDFS using a file transfer protocol (FTP) batch. User profile data from the business support system (BSS) are also stored in HDFS.

We used Kafka (a distributed messaging system "developed for collecting and delivering high volumes of log data with low latency" [28]) and Spark (a cluster computing technology utilizing in-memory cluster computing to increase processing speed [29]) for converting the CDR log and user profile data into cross-CDRs. To process several billion records in real-time, we allocate 12 servers with 384 cores to the conversion. The six Kafka servers collect CDRs from HDFS and feed them to twelve Spark servers. The Spark servers process CDRs to generate new records in the cross-CDR database. As a single call has multiple CDRs (CDR-S1, CDR-SIP, and CDR-VoLTE), they are combined into one record. The Spark composer module identifies CDRs from the identical calls using phone number and call start/end time and allocates a call ID. It also appends KPIs for connections, KPIs for SIP (session initiation protocol is a signaling protocol for multimedia sessions [30]), and KPIs for VoLTE and equipment ID. The equipment information and customer profiles are also added to the cross-CDRs by the Spark servers. The data in the cross-CDR database flows into the perceived quality model to evaluate the QoS at the



Phone number	Call ID	Start/end time	S1 KPIs	SIP KPIs	VoLTE KPIs	Device information	User profile	Occupied equipment ID
010-XXXX-YYYY	1122	20160822010311: 20160822010256	Attach attempt Attach success Attach fail	SIP registration attempt SIP registration success SIP registration fail	eMOS defect rate Packet loss rate Delay defect rate Jitter defect rate Drop rate ...	Device name Device vendor OS name OS version Firmware version ...	Sex Age Registration date Charging plan ...	Cell ID DU ID Router ID MME ID SPGW ID HSS ID CSCF ID ...
010-YYYY-XXXX	1123	20160822010320: 20160822010259	Paging attempt Paging success Paging fail ...	Call attempt Call connect Call success ...				

eNB: evolved Node B
MME: mobility management element
SGW: serving gateway

PGW: packet gateway
CSCF: call session control function
HDFS: Hadoop distributed file system

SIP: session initiation protocol
BSS: business support system
eMOS: E-model/MOS

FIGURE 1: System diagram and an example cross-CDR database.

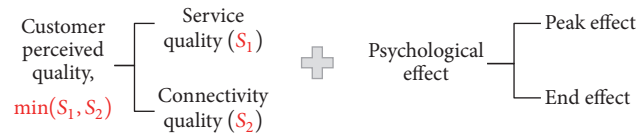


FIGURE 2: Customer experience quality structure.

call level in real-time; the customer-perceived quality score is generated at the subscriber level.

4. Developing a Perceived Quality Score Model for Complaint Calls

In this section, we discuss our score model to predict customer complaint calls, because the complaint calls are a proxy for customer-perceived quality. Our work is focused on voice call quality. To develop the voice call quality model, we work with call quality-related customer complaints only, excluding the other ones such as billing, device, and so on. As shown in Figure 2, the score model consists of service quality and connectivity quality. We develop two models S_1 and S_2 for connectivity quality and service quality, respectively. These

scores are combined with the final experience score S in the following manner:

$$S = \min(S_1, S_2). \quad (1)$$

Each S_i , $i = 1$ and 2 , ranges from 1 to 100. We use the min function $\min(S_1, S_2)$, supposing that the worse of the two scores, S_1 and S_2 , will determine the customer experience. We also consider the psychological effect of *peak* and *end* effects. The peak and end effect implies that people judge an experience largely based on how they felt at its peak and at its end, rather than based on the sum or average of the individual moments of the experience. To capture the worst experience, we used a SPC-based approach, which is described in Section 4.3(2).

Development of a quantitative customer-perceived quality model is useful for improving the a priori customer satisfaction. Improved service and complaint handling are possible if the perceived quality levels are better understood by service providers. However, utilizing the complaint call as a target variable is challenging for several reasons.

4.1. Challenges

(1) *Customers with No Action*. It is well-known that a large portion of dissatisfied customers take no action [10]: for

TABLE 1: Variable selection via clustering analysis.

Group	Group characteristics	Size (%)	Comp. call ratio (normal)	Drop rate	TX def. rate	RX def. rate
(1)	Normal	95.4	0.08% (1.00)	0.09%	0.50%	1.75%
(2)	Bad one-way call	4.5	0.12% (1.57)	0.16%	0.12%	4.87%
(3)	Bad drop rate, bad eMOS	.01	0.30% (3.75)	70.52%	9.46%	2.81%

services, 29.4% of customers with low quality took no action. On the other hand, for telecommunication services, customers with good network quality sometimes also make complaint calls. The combination of these two opposite errors makes the development of a good scoring system a challenging problem.

(2) *Imbalanced Data*. Learning from imbalanced datasets can be problematic, making meaningful analyses difficult. As complaint calls account for a very small fraction (less than 0.01% of daily calls), it is difficult to create an accurate model for predicting complaint calls. We have tried the known methods to tackle the imbalance of the data such as the undersampling, and so on [31, 32], without success (the following methods were tried [31]: (1) collection of additional data, (2) resampling, and (3) different algorithms. We first tested with the data from the past three months without success. The undersampling method was then tried by matching the ratio of the two user groups (those who made calls and those who did not make complaint calls). We then applied various methods including random forest, SVM, LASSO, and decision tree. Those methods exhibit an accuracy ratio between 59.9% and 72.6%, while the accuracy of our method is 91.5%).

(3) *Skewed Independent Variables*. Many KPIs (such as call drop rate or blocking rate), which are used as explanatory variables, have very biased distributions ranging from 0.1% to 0.01% or less. Therefore, these indices are not good candidates for independent variables.

4.2. Key Ideas

(1) *Independent Variable Selection with Clustering*. Typical supervised learning methods such as regression and the support vector machine did not generate meaningful models because of the imbalanced data and/or the skewness of the variables. We repeated an unsupervised method of cluster analysis until we found a cluster with above-normal complaint call rates. We analyzed and characterized these clusters to choose meaningful variables to use as independent variables.

(2) *SPC-Based Poor Quality Pattern Variable Selection*. The “peak and end rule” in psychology states that people judge an experience largely based on how they felt at its peak and at its end, rather than based on its average [33]. To model an abnormally poor experience of a voice call, we use statistical process control (SPC), which is useful for detecting abnormal patterns. Among many rules of the SPC, we selected a few of them that significantly explain the likelihood of complaint calls.

(3) *Use of Pseudo Target Variable via PCA*. We do not use a linear regression model with the target variable of complaint calls with, since it is difficult to estimate individual customers’ call quality experiences. Instead, we construct a pseudo variable representing the independent variables; this pseudo variable will be used as a target variable. We set the pseudo variable as the first principal component in principal component analysis (PCA). The pseudo variable is used to maintain the characteristics of the independent variables while relaxing the imbalanced class characteristics of the target variable.

4.3. *Service Quality Score S_1* . The service quality score S_1 is a model to explain customer-perceived call quality (complaint calls) during service. We apply the four steps shown in Figure 3: (1) variable selection via clustering; (2) SPC-based abnormal quality pattern variable selection; (3) pseudo target Y_p via PCA; and (4) final model selection.

(1) *Variable Selection via Clustering*. First, we identify the 16 candidate variables among the KPIs of the quality of calls. These variables are related to packet loss rate, delay, jitter, and bursts. After cleansing the data, we repeated K -means analysis by varying the number of clusters and weights of each variable until we find a meaningful cluster with much higher complaint call rates than the average. Table 1 shows the clustering results for three groups, one normal group (1) and two poor groups (2 and 3). The size of the normal group is 95.4%; that of the poor groups is 4.6%. Note that the second and third groups have 1.57 and 3.75 times higher complaint calls rates, respectively, than that of the normal group.

From the results of the clustering analysis, the variables that significantly explain the complaint call rate are identified. We dropped several variables with low explanatory power considering the multicollinearity between the variables and each variable’s explanation rate. We chose the drop defect rate (x_1), Tx eMOS defect rate (x_2) (the eMOS stands for the E-model/MOS score [18]), and Rx eMOS defect rate (x_3) as the independent variables for the customer-perceived quality model, where the eMOS score ranges from 1 (worst) to 5 (best). The drop defect rate (x_1) measures the drop rate of the calling services. The eMOS score is calculated every 5 seconds, and it is considered as defect if the score is less than or equal to 2. The eMOS defect rate is the ratio of the number of eMOS defect to the total number of eMOS scores. Tx eMOS defect rate (x_2) and Rx eMOS defect rate (x_3) represent the fraction of poor quality period for outgoing and incoming calls, respectively.

(2) *SPC-Based Abnormal Quality Pattern Variables Selection*. Based on the peak and end rule from the field

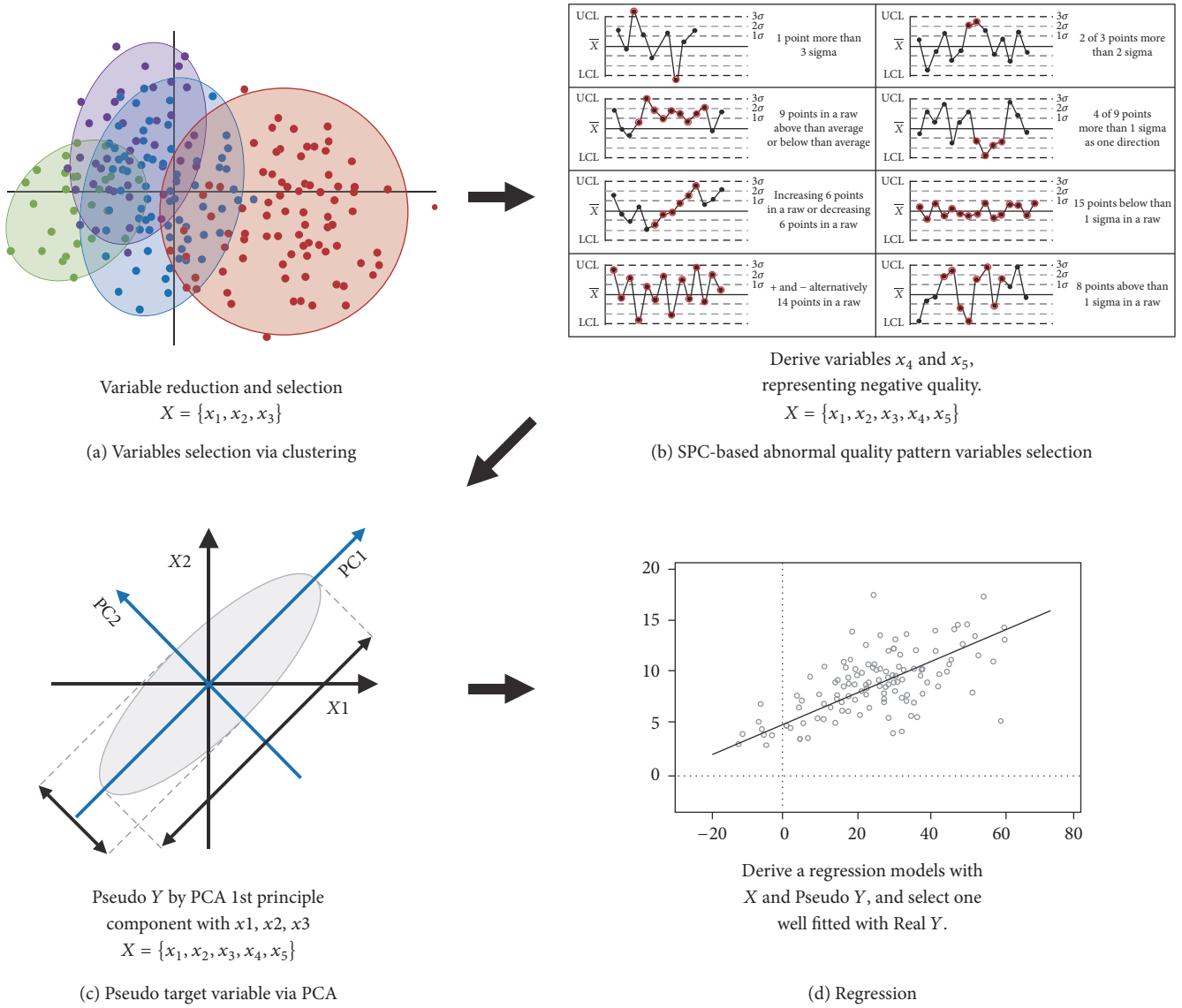


FIGURE 3: Procedure of service quality modeling.

of psychology, we try to detect abnormally poor QoS as it can affect the perceived quality for customers. We adopted a method based on SPC, which is known to be very helpful for detecting abnormal behavior [34]. We used popular rules called the Western electric rules (for the zone rules and the asymmetric rules, refer to https://en.wikipedia.org/wiki/Western_Electric_rules), especially the zone the asymmetric rules, which are used to find abnormal patterns in SPC (for other approaches to detect anomalies, refer to [35, 36]), for example, “two consecutive values higher (or lower) than 2σ ” or “four consecutive points out of five higher (or lower) than σ ” to detect the bad perceived quality from frequent, consecutive, or intense defects at the subscriber level.

Each call generates a sequence of eMOS scores, as shown in Figure 4. Each point in the figure corresponds to an eMOS score of $t = 5$ seconds. The \bar{X} is the mean of eMOS scores. In

Rule 1, if two consecutive points are higher (or lower) than 2σ , we judge it as an anomaly. As “out of 2σ ” is a rare event, it shall be even rarer to observe its continuous appearance as in Rule 1.

Among the candidates in Table 2, Rules 1 and 2, are selected as variables x_4 and x_5 , respectively. If the ratio of A to B is higher than a threshold (we use 1.5 as the threshold. There is a significant gap observed between the ratios (A/B) of Rules 2 (1.625) and 3 (1.083)), we consider the rule to have correlation with customer complaint calls where A is the average complaint call rate for those who exhibit the abnormal pattern of rules, and B is the rate for those who have no such patterns. The variable x_4 counts the number of patterns of “2 consecutive points higher (or lower) than 2σ ” from the chart. Similarly, the variable x_5 counts the number of patterns “4 consecutive points out of five higher (or lower) than σ ”.

TABLE 2: Selection of rules for derived variables.

Pattern	Complaint call rate	Complaint call rate	Ratio (A/B)
	Matching pattern (A)	No pattern (B)	
Rule 1	2 consecutive points higher (or lower) than 2σ	0.27	1.963
Rule 2	4 consecutive points out of five higher (or lower) than σ	0.08	1.625
Rule 3	1 point higher (or lower) than 3σ	0.12	1.083
Rule 4	6 consecutive increasing/decreasing points	0.16	1.063
Rule 5	8 consecutive points out of $\pm 1\sigma$	0.32	0.969
Rule 6	9 points higher (or lower) than average	0.27	0.926
Rule 7	14 consecutive alternating points	0.32	0.906
Rule 8	15 consecutive points out of $\pm 1\sigma$	0.21	0.857

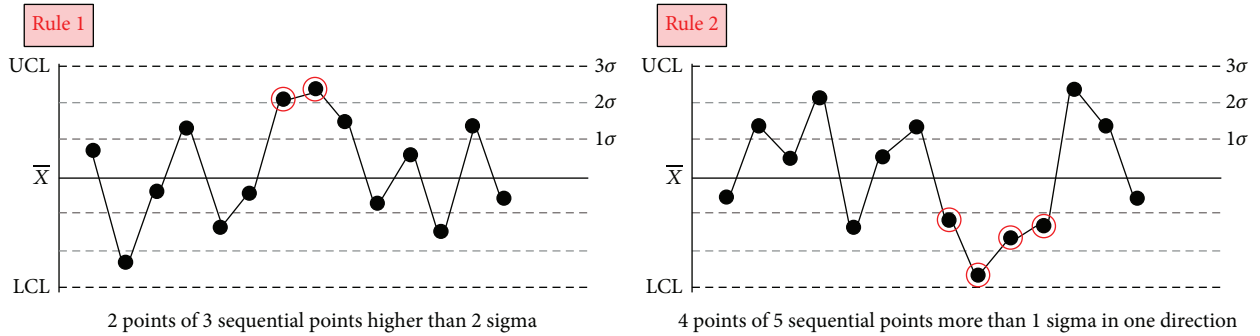


FIGURE 4: Matched SPC rule having an effect on the complaint call rate.

(3) *Pseudo Y via PCA*. PCA is commonly used to extract the representative characteristics of the dataset. Because of the imbalanced class characteristics of the variables, we could not develop a meaningful regression model for Y as a complaint call rate. Instead, we generate the pseudo variable Y_p by applying PCA to sets of x_1, x_2 , and x_3 . We set the first principal component as the pseudo target variable Y_p . We did not include the two derived variables, x_4 and x_5 , for the PCA, so as to avoid distortion caused by the correlated independent variables.

We constructed regression models using the five independent variables (including the two derived variables) to explain the pseudo variable using various construction options. Then, we choose a regression model whose estimations show the greatest similarity with the values of the true target variable as the final call quality model. We tested the model's goodness of fit using the complaint call rate, not using statistical performance indexes, so as to increase the explanatory power of our regression model. We calculated the average complaint call rate of users whose score is less than or equal to 30 and selected a model with the highest complaint call rates.

(4) *Service Score Model S_1* . The resultant service regression model Y_1 is given by

$$Y_1 = 0.363x_1 + 0.526x_2 + 0.132x_3 + 2.86x_4 + 0.70x_5, \quad (2)$$

where x_1 is the drop defect rate, x_2 is the TX eMOS defect rate, x_3 is the Rx eMOS defect rate, x_4 is the count of Rule 1, and x_5 is the count of Rule 2. The variable Y_1 is zero if all variables are zero, which means best quality; a higher Y_1 means a poorer quality for customers.

TABLE 3: Scoring of objective quality score S_1 .

Score S_1	0	10	20	30	40	50	60	70	80	90	100
Percentile	100.0	99.0	95.0	90.0	75.0	62.5	50.0	37.5	25.0	12.5	0.0

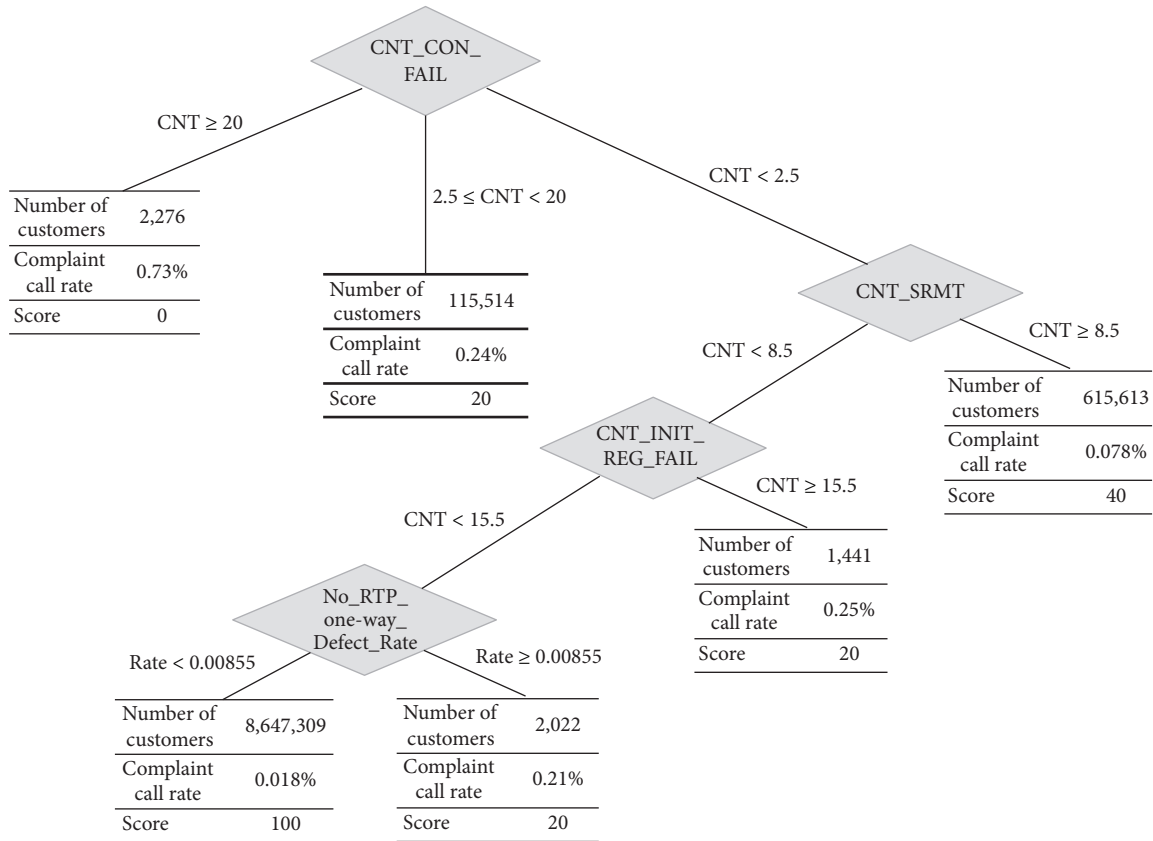
Based on the value of Y_1 , the output of the regression model, the score S_1 is calculated as shown in Table 3. The zero-percentile ($Y_1 = 0$) (the zero-percentile corresponds to the regression model value Y_1 of 0.0) gets the score of 100, and 99-percentile ($Y_1 = 0.5$) (the 99 percentile corresponds to the regression model value Y_1 of 0.5) gets the score of 10. The idea behind the function is that we would like to focus on 10% of customers who are highly likely to make complaint calls. Those are who have scores less than 30. In those regime, 1-percentile has score less than 10; 5-percentile has score less than 20; and 10-percentile has score less than 30. For the other 90 percent of customers, we use a linear score function $S_1 = 100 - 0.8P$, where P is the percentile except scores between 30 and 40.

4.4. *Connectivity Quality Score Model S_2* . The connectivity score is developed to model the perceived quality of connection establishment and connection drop. To evaluate the connectivity quality score, we selected approximately 100 KPI variables from SI/SIP/VoLTE CDRs, some of which are shown in Figure 1. With basic statistical analyses, we chose 24 variables, as shown in Table 4, that can represent the overall connectivity quality: we omitted variables with missing values and high correlations using the multicollinearity test.

TABLE 4: Selection of variables for connectivity quality.

Reference CDR	Selected KPIs for connectivity quality
VoLTE CDR (8)	Drop count/connect success rate/connect fail rate/drop rate/no Tx-RTP one-way-call defect/no Rx-RTP one-way-call defect/no Tx-RTP one-way-call defect rate/no Rx-RTP one-way-call defect rate
SIP CDR (4)	Initial-REGISTER fail count/initial-REGISTER success rate/periodic-REGISTER fail count/periodic-REGISTER success rate
SI CDR (12)	Attach fail count/attach success rate/SRMO fail count/SRMO success rate/SRMT fail count/SRMT success rate/SIHO/TAU(A)/TAU(I) fail count/SIHO/TAU(A)/TAU(I) success rate

SRMO: service request mobile origination; SRMT: service request mobile termination; SIHO: SI handover; TAU: tracking area update.



CNT_CON_FAIL: the counts on failed connections

CNT_SRMT: the counts on service request mobile termination

CNT_INIT_REG_FAIL: the counts on (SIP) initial registration fail

No RTP one-way defect rate: the defection rate that there

is no RTP packet, or there is

only one-way RTP packet

FIGURE 5: Decision tree model for connectivity quality.

Except the decision tree model, most predictive models with supervised learning did not work because of the high imbalance of datasets and skewness of variables (unlike service quality models, we think that a decision tree model is obtainable because the connection quality is either one (if a service is connected) or zero (otherwise)). We used SAS E-miner package [37] to derive the decision tree with the C4.5 algorithm [38] to find the decision tree model for complaint calls of Figure 5. The target variable is whether a call contains complaint calls or not; the independent variables are those 24 in Table 4. We varied the parameters (such as maximum number of branches, maximum depth, and leaf size) to

find the decision tree with the highest accuracy. The tree showed the performances of the accuracy of 99.6146% and the sensitivity of 83.3333%, and the specificity of 99.6151%.

The decision tree shows that the number of connection failures (CNT_CON_FAIL) is very important. If this number is higher than 20, the score is zero; if it is between 2.5 and 20, the score is 20. The next branch is the number of service request mobile termination (SRMT) failures (CNT_SRMT_FAIL). After a connection has been established, if the service request fails, CNT_SRMT_FAIL is increased by one. If the number of failures is higher than 8.5, the score is 40; otherwise, the next variable is the number of

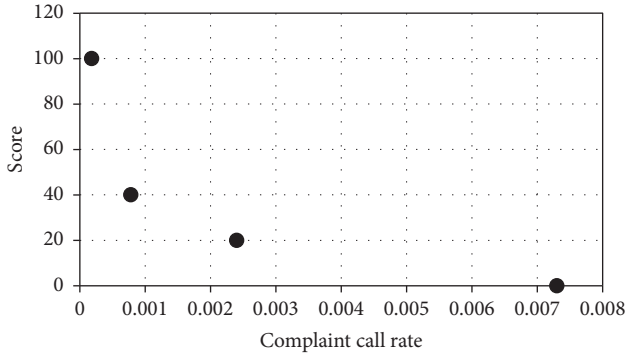


FIGURE 6: Connectivity quality score versus complaints call rate.

initial registration failures (CNT_INIT_REG_FAIL), which happens in the authentication phase. If the number of failures is higher than 15.5, the score is 20. Otherwise, the method checks the real-time transport protocol (RTP) defect rate (RTP_DEF_RATE). The RTP packets should be shown in both directions. However, if they are shown in only one way or less, it can be considered as a defect. The RTP defect rate is the number of RTP defect to the total number of connections. If the rate is higher than the threshold, the score is 20. Otherwise, the score is 100.

The connectivity score S_2 is calculated using the percentile-score mapping of Table 3. Each box in the decision tree of Figure 5 has a complaint call rate. For example, the complaint call rate of the bold-line box under the CNT_CON_FAIL node in Figure 5 is 0.24%. The complaint call rate is mapped into percentile; the percentile is mapped into the score using Table 3. The rate 0.24% corresponds to 95 percentile and the score of 20.

Figure 6 shows the complaint call rates versus connectivity quality score. As the decision tree generates 4 scores of 0, 20, 40, and 100, there are 4 dots in the Figure. It shows that the extreme poor quality score of 0 has high complaint call rates 0.73%, while the score of 100 has that of 0.018%. The worst score group has 40 times higher rate than the best score group.

Perceived Quality Score. The perceived quality score S is the minimum of the two scores, namely, service score S_1 and connectivity score S_2 , as Equation (1). This choice can be justified from the fact that customers react more to worse experiences (minimum of two) than average experiences (weighted sum) by the peak and end rule.

5. Experiments

We developed the model using the data collected from August 10 to 16, 2016, from a production network. It contains CDRs of 35 billion calls from 9 million subscribers. The validation is performed with the test data from August 17 to 23, 2016.

5.1. Perceived Quality Score versus Complaint Call Rate. Figure 7 shows the complaint call rates versus perceived quality score $\min(S_1, S_2)$. The line with diamonds is drawn

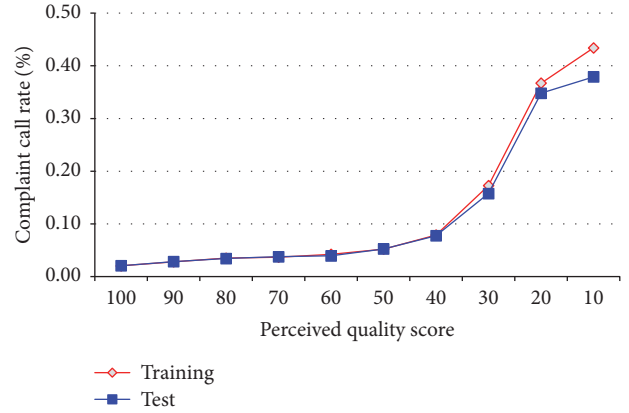


FIGURE 7: Perceived quality score versus complaints call rate graph.

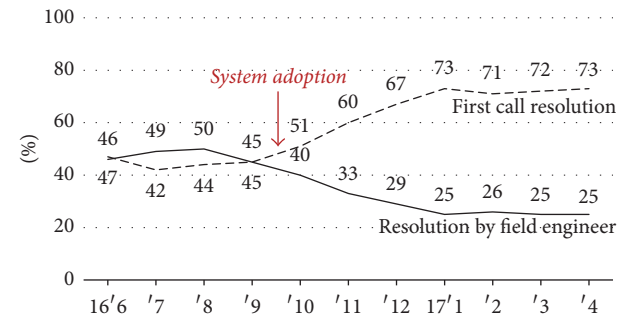


FIGURE 8: System performance before and after: the first call resolution rate increases and the field dispatch rate decreases.

using the training data, while the other line with squares results from the validation data. The two lines show very similar pattern with the largest difference of 11% at the score of 10. The complaint call rate is low when the objective score is higher than or equal to 40 and starts jumping from the score of 30 or less. The average complaint call rate of the low score group (30 or less) is six times higher than that of the high score group (higher than 30) and it is 11 times higher than the average. The complaint call ratio of the worst group (score 10 or less) to that of the best score group (higher than 90) is 22. The figure confirms the validity of the proposed score as a good measure predicting complaint calls.

5.2. Comparison with Random Guess Model. Table 5 shows the performance comparison of our model with that of a random guess model, which does random coin flipping for the prediction. The random guess model does coin flipping with probability of 0.067% for complaint call, which is the average complaint calls.

The true positive value, $1.08e - 6$, of the proposed model is 2.4 times higher than $0.45e - 6$ of the random guess model. Observe also that true negative value is improved by a similar amount even though it is very small due to high imbalance of the data. Observe also that the accuracy, sensitivity, and specificity are also improved.

5.3. Comparison of KPI versus CEI-Based Model: Investment Targets. We selected the top n worst performing cells, which

TABLE 5: Performance comparison with random guess model.

	Score model (A)	Random model (B)	A/B	A - B
True positive	$1.08e - 6$	$0.45e - 6$	2.41	$0.64e - 6$
True negative	99.86571%	99.86565%	1.000001	$0.64e - 6$
Accuracy	99.86582%	99.86569%	1.000001	$1.28e - 6$
Sensitivity	0.16208%	0.06720%	2.41	$9.48e - 3$
Specificity	99.93286%	99.93280%	1.000001	$0.64e - 6$

TABLE 6: Comparison of complaint call rates for selected n base stations for $n = 10, 50, 100,$ and 200 .

Selected n cells	CEI-based (A)	KPI-based (B)	A/B
10	0.78%	0.13%	6.0
50	0.43%	0.08%	5.3
100	0.32%	0.62%	5.2
200	0.11%	0.05%	2.1

are top candidates for upgrade and repair, from both methodologies and compared them as shown in Table 6. In the KPI-based method, which is the method used currently, important KPIs such as the cell outage rate, the paging success rate, the number of the radio link failures, the number of radio resource control setup failures, and so on are considered. Each KPI value is mapped into a KPI score between 0 and 100, and the total score of the cell is a weighted sum of the KPI scores. For example, the more the outage rate, the lower the score. In the new CEI-based model, we defined *bad quality users* as those who have score less than or equal to 30. We counted the number of *bad quality users* in all cells and sorted cells in a descending order of the number of *bad quality users*.

It compares the average complaint call rates from the selected n cells. Observe that the complaint call rates from the CEI-based methods are two to six times higher than those from the KPI-based method. The average complaint calls from top 10 cells are 0.78% and 0.13%, respectively. The CEI-based method is able to select cells that cause complaint calls better than the KPI-based method.

5.4. Complaint Call Response Performance. Figure 8 shows the changes in the first call resolution (FCR) rate in the call center and the field dispatch rate trends after the adoption of the system. The FCR rate is the fraction of customers whose issues were resolved the first time they called. We monitored the trends up to seven months after adoption. Prior to the introduction of the system, the FCR rate was about 45%. After adoption, the FCR rate increased gradually to 51%, 60%, 67%, and 73%, and then leveled off at around 72%. The decreasing line shows the field dispatch rate (FDR) of field engineers. When issues are not resolved at the call center, the field engineers are dispatched to solve them. Prior to adoption, the FDR was between 45% and 50%. After the adoption, it decreased to 25%, which is about half of the prior rate. With the adoption of the system, the call center representative can more accurately determine the customer quality issues and resolve many of them in advance, which reduces the demands on the field engineers.

6. Conclusions

Using a machine learning methodology, we developed a scoring model to predict the likelihood of customer complaint calls as a proxy for customer-perceived quality. The developed model consists of two components: service quality and connectivity quality. The service quality model considers the perceived service quality during the sessions, while the connectivity quality considers connection establishment. We overcame several issues using clustering-based variable selection, pseudo target variable generation using PCA, and the introduction of SPC-based variables. We implemented the model in a real production network system that handles several billion calls per day. To manage the tens of billions of CDRs, we used the open source Kafka and Spark software packages. The developed system generates a cross-CDR database to be used by the quality model after merging various CDRs.

The validation test showed that the score model has strong explanatory power. Individuals from the lowest scored group (a score less than or equal to 30) have a 20 times higher likelihood of making complaint calls than the highest score group. Also, the sensitivity of the proposed model is 2.4 times higher than that of the random guess model. Compared to the legacy system, the new system based on the proposed scoring model detects base stations that cause a high level of complaint calls. When cells are selected with the new method, the complaint calls made in the top ten worst performing cells increase by about six times. By upgrading or repairing these cells first, we believe that customer dissatisfaction can be handled more efficiently. When we monitored the first call resolution rate, that is, the fraction of customers whose issues are resolved at the first call, the rate increased from 45% to 72% within six months of adoption, which shows the effectiveness of the proposed system.

Integration of a subjective quality model with the overall quality model is a future research direction. Additionally, the differences in the data service quality or video service quality from the call quality can be integrated with the objective quality model. Our final goal is to generate data that can be used to improve business performance by analyzing the influence of network quality on customer net profit scores and customer churn.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

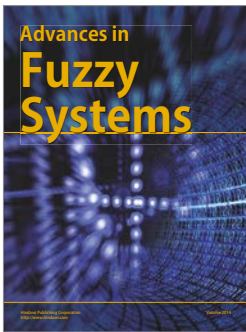
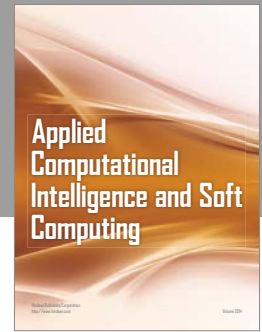
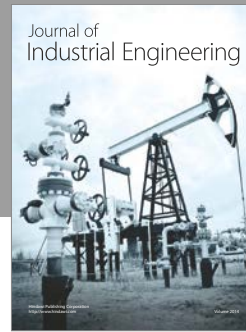
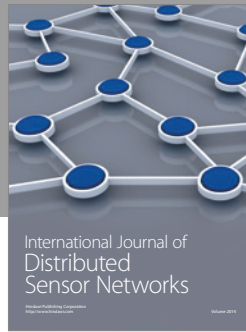
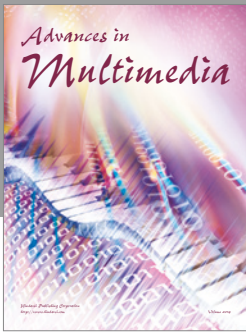
Acknowledgments

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2012R1A1A2007890).

References

- [1] J.-H. Ahn, S.-P. Han, and Y.-S. Lee, "Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry," *Telecommunications Policy*, vol. 30, no. 10-11, pp. 552–568, 2006.
- [2] R. Siber, "Combating the churn phenomenon," *Telecommunications (International Edition)*, vol. 31, no. 10, pp. 77–81, 1997.
- [3] R. Schatz, T. Hoßfeld, L. Janowski, and S. Egger, "From packets to people: quality of experience as a new measurement challenge," in *Data Traffic Monitoring and Analysis*, vol. 7754 of *Lecture Notes in Computer Science*, pp. 219–263, Springer, Berlin, Germany, 2013.
- [4] N. Khan, I. Yaqoob, I. A. T. Hashem et al., "Big data: survey, technologies, opportunities, and challenges," *Scientific World Journal*, vol. 2014, Article ID 712826, 18 pages, 2014.
- [5] K. A.-S. Woo and H. K. Y. Fock, "Customer satisfaction in the Hong Kong mobile phone industry," *Service Industries Journal*, vol. 19, no. 3, pp. 162–174, 1999.
- [6] J. Lee, J. Lee, and L. Feick, "The impact of switching costs on the customer satisfaction-loyalty link: Mobile phone service in France," *Journal of Services Marketing*, vol. 15, no. 1, pp. 35–48, 2001.
- [7] D. Finkelman, R. Cetlin, and D. Wenner, "Marketing customer satisfaction efforts pay off—part A," *Telephony*, pp. 20–24, 1992.
- [8] D. Finkelman, R. Cetlin, and D. Wenner, "Marketing customer satisfaction efforts pay off—part B," *Telephony*, pp. 24–32, 1992.
- [9] A. D. Athanassopoulos and A. Iliakopoulos, "Modeling customer satisfaction in telecommunications: Assessing the effects of multiple transaction points on the perceived overall performance of the provider," *Production and Operations Management*, vol. 12, no. 2, pp. 224–245, 2003.
- [10] R. L. Day and E. L. Landon Jr., *Toward a Theory of Consumer Complaining Behavior*, North-Holland, New York, NY, USA, 1977.
- [11] B. Tronvoll, "A dynamic model of customer complaining behaviour from the perspective of service-dominant logic," *European Journal of Marketing*, vol. 46, no. 1, pp. 284–305, 2012.
- [12] J. Turkka, F. Chernogorov, K. Brigatti, T. Ristaniemi, and J. Lempiäinen, "An approach for network outage detection from drive-testing databases," *Journal of Computer Networks and Communications*, vol. 2012, Article ID 163184, 13 pages, 2012.
- [13] H. Li, D. Wu, G.-X. Li et al., "Enhancing telco service quality with big data enabled churn analysis: infrastructure, model, and deployment," *Journal of Computer Science and Technology*, vol. 30, no. 6, pp. 1201–1214, 2015.
- [14] J. Moysen, L. Giupponi, and J. Mangues-Bafalluy, "A mobile network planning tool based on data analytics," *Mobile Information Systems*, vol. 2017, Article ID 6740585, 16 pages, 2017.
- [15] Y. Yin, W. Xu, Y. Xu, H. Li, and L. Yu, "Collaborative QoS prediction for mobile service with data filtering and SlopeOne model," *Mobile Information Systems*, vol. 2017, Article ID 7356213, 14 pages, 2017.
- [16] S. Peng, J. O. Fajardo, P. S. Khodashenas et al., "QoE-oriented mobile edge service management leveraging SDN and NFV," *Mobile Information Systems*, vol. 2017, Article ID 3961689, 14 pages, 2017.
- [17] ITU-T Recommendation P.800, "Methods for subjective determination of transmission quality," 1998.
- [18] ITU-T Recommendation G.107, *The E-model: a computational model for use in transmission planning*, 2009.
- [19] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," 2008.
- [20] C. Yurtseven, "Complaint calls as a proxy for perceived quality: the turkish dishwasher demand estimation," *Journal of Global Analysis*, vol. 6, no. 1, pp. 71–84, 2016.
- [21] J. Niemöller and N. Washington, "Subjective perception scoring: psychological interpretation of network usage metrics in order to predict user satisfaction," in *Proceedings of the 19th International Innovations in Clouds, Internet and Networks*, 2016.
- [22] C. Fornell and B. Wernerfelt, "Defensive Marketing Strategy by Customer Complaint Management: A Theoretical Analysis," *Journal of Marketing Research*, vol. 24, no. 4, p. 337, 1987.
- [23] J. Laiho, K. Raivio, P. Lehtimäki, K. Hatonen, and O. Simula, "Advanced Analysis Methods for 3G Cellular Networks," Report A65, Publications in Computer and Information Science, Helsinki University of Technology, 2002.
- [24] M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," *IEEE Network*, vol. 24, no. 2, pp. 36–41, 2010.
- [25] "Qualinet white paper on definitions of quality of experience," in *European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Version 1.2*, P. Le Callet, S. Möller, and A. Perkis, Eds., 2013.
- [26] TM Forum Best Practices RN341, "Customer Experience Management Index (CEMI)," Available: <https://www.tmforum.org/resources/standard/tr193-customer-experience-management-index-v1-3/>.
- [27] "An Introduction to LTE," 3GPP LTE Encyclopedia. Retrieved December 3, 2010. <https://sites.google.com/site/lteencyclopedia/home>.
- [28] J. Kreps, N. Narkhede, and J. Rao, "Kafka: a distributed messaging system for log processing," in *Proceedings of the 6th International Workshop on Networking Meets Databases (NetDB)*, 2011.
- [29] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica, "Spark: cluster computing with working sets," in *Proceedings of the HotCloud'10*, 2010.
- [30] RFC 3261. SIP: Session Initiation Protocol. IETF. 2002.
- [31] J. Brownlee, "8 tactics to combat imbalanced classes in your machine learning dataset," *Machine Learning Process*, 2015, <http://machinelearningmastery.com/tactics-to-combat-imbalanced-classes-in-your-machine-learning-dataset/>.
- [32] N. V. Chawla, "Data mining for imbalanced datasets: an overview," in *Data Mining And Knowledge Discovery Handbook*, pp. 875–886, Springer, 2009.
- [33] D. Kahneman and A. Tversky, "Experienced utility and objective happiness: a moment-based approach," *The Psychology of Economic Decisions*, vol. 1, pp. 187–208, 2003.
- [34] D. C. Montgomery, *Introduction to Statistical Quality Control*, John Wiley & Sons, 2007.
- [35] P. Casas, P. Fiadino, S. Wassermann et al., "Unveiling network and service performance degradation in the wild with mplane," *IEEE Communications Magazine*, vol. 54, no. 3, pp. 71–79, 2016.

- [36] P. Casas, A. D'Alconzo, T. Zseby, and M. Mellia, "Big-DAMA: big data analytics for network traffic monitoring and analysis," in *Proceedings of the the 2016 Workshop LANCOMM@ SIGCOMM*, pp. 1–3, Florianopolis, Brazil, August 2016.
- [37] "SAS E-miner," Available: https://www.sas.com/en_us/software/enterprise-miner.html.
- [38] S. Ruggieri, "Efficient C4.5 [classification algorithm]," *IEEE Transactions on Knowledge and Data Engineering*, vol. 14, no. 2, pp. 438–444, 2002.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

