Machine learning techniques for transmission parameters classification in multi-agent managed network

1st Dariusz Żelasko
Department of Information and Communications Technology, Faculty of Computer Science and Telecommunications, Cracow University of Technology
Krakow, Poland
dzelasko@pk.edu.pl
ORCID: 0000-0001-5125-3647

2nd Paweł Pławiak
Department of Information and Communications Technology, Faculty of Computer Science and Telecommunications, Cracow University of Technology
Krakow, Poland
Institute of Theoretical and Applied Informatics, Polish Academy of Sciences
Gliwice, Poland
plawiak@pk.edu.pl or plawiak@iitis.pl
ORCID: 0000-0002-4317-2801

3rd Joanna Kołodziej
Research and Academic Computer Network - National Research Institute (NASK)
Warsaw, Poland
joanna.kolodziej@nask.pl
ORCID: 0000-0002-5181-8713

ABSTRACT

Looking at the rapid development of computer networks, it can be said that the transmission quality assurance is very important issue. In the past there were attempts to implement Quality of Service (QoS) techniques when using various network technologies. However QoS parameters are not always assured. This paper presents a novel concept of transmission quality determination based on Machine Learning (ML) methods. Transmission quality is determined by four parameters - delay, jitter, bandwidth and packet loss ratio. The concept of transmission quality assured network proposed by Pay&Require was presented as a novel multi-agent approach for QoS based computer networks. In this concept the essential part is transmission quality rating which is done based on transmission parameters by ML techniques. Data set was obtained based on the experience of the users test group. For our research we designed a machine learning system for transmission quality assessment. We obtained promising results using four classifiers: Nu-Support Vector Classifier (Nu-SVC), C-Support Vector Classifier (C-SVC), Random Forest Classifier, and K-Nearest Neighbors (kNN) algorithm. Classification results for different methods are presented together with confusion matrices. The best result, 87% sensitivity (overall accuracy), for the test set of data, was achieved by Nu-SVC and Random Forest (13/100 incorrect classifications).

Keywords

multi-agent system, QoS, Pay&Require, resource allocation, economical models, machine learning

I. INTRODUCTION

Over the past years we can observe fast evolution of computer networks. New services need communication with various devices, which makes the popular transmission protocols and devices, such as 56k modems, unuseful. Generally computer networks work with best effort behavior - without guaranteed transmission quality. Such method of operation seems to be insufficient if we look at the modern use of computer networks. Data transmission that meets customer requirements is becoming increasingly important. The Pay&Require [1], [2] was proposed to achieve this target. In Pay&Require transmission quality is determined by delay, jitter, bandwidth and packet loss ratio. Combination of these parameters should provide clear information about transmission quality. Such classification is not easy. The motivation of this work is to indicate the possibility of using machine learning to determine the quality of transmission based on certain parameters.

Our contributions are as follows:

- collecting a data set based on users experience,
- a novel idea of using the ML methods for transmission quality classification in computer networks,
- the application of a novel method based on genetically optimized classifiers coupled with cross-validation [27] and feature selection.

978-1-7281-6095-5/20/$31.00 ©2020 IEEE
DOI 10.1109/CCGrid49817.2020.00-20

Authorized licensed use limited to: Politechnika Krakowska. Downloaded on September 01,2020 at 09:55:01 UTC from IEEE Xplore. Restrictions apply.

Data set was obtained based on the experience of the users test group. Experience means, in this case, empirical perception of transmission quality. Four users were tested, each user rated 100 samples. The sample was a video stream and a website - displayed with varying quality of transmission parameters. Afterwards we designed a machine learning system for transmission quality assessment. This system consisted of 5 stages based on: data preprocessing (5 types), feature selection, cross-validation (2 types), designing ML algorithms (4 types) and parameter optimization. The research brings a new methodology for classifying transmission quality.

The rest of the paper is organized as follows. In Sec.II we present a state–of–the–art in the domain. The concept and the details of the modules and data of the designed system are presented in Sec.III. In Sec.IV we provide the experimental analysis and discuss the results. The paper ends in Sec.VI with simple conclusions and plan for the future work.

II. Related Work

There are different network techniques e.g. ATM (asynchronous transfer mode), MPLS (multiprotocol label switching), GMPLS (generalised MPLS) or SDN (software-defined networking) which provides Quality of Service (QoS) techniques that include the ability to ensure transmission parameters. In general, mentioned technologies can be classified as centralized. SDN is an example of programmable network [3]. In this concept network is controlled and managed dynamically by open interfaces. SDN is based on approach that separates data forwarding from logic that controls it. There is central controller which communicates with physical plane (devices) providing them necessary information. QoS is implemented in control plane which is used to monitor and define control parameters. Pay&Require concept of quality assured network technique, as novel decentralized system was proposed and presented in Sec.II-A.

Machine learning seems to be very popular in different areas [13]–[15], [17], [18], [20]–[23] but in computer networks it is still not very popular. Works were carried out regarding use of machine learning in computer networks. In [8] authors have done a review of works within period 2004 - 2007 typing about network traffic classification with use of machine learning. The general problem is that normally traffic classification is done based on TCP or UDP port numbers or contents of packet payloads. Such method is not reliable because user can use different techniques to avoid filtering based on port number. Authors made deep research in usability of machine learning in IP traffic classification. The conclusion of this paper is that such techniques can be used. Problematic might be real-time behaviour. Quickly and accurately traffic classification is very important for QoS and security, especially for unknown traffic flows. In [5] authors proposed the solution based on SVM (Support Vector Machine) method [11], [16]. They have used SVM to train 7 classes of traffic. Proposed solution works in real-time, it checks headers of the packets. The main problem of SVM method is the expectation of a large number of labeled training samples, also classification based on whole flow might be too late so necessary techniques should be applied at early transmission stage [6]. In [7] SVM use in CoMo architecture was proposed. CoMo provides software abstraction layer for real-time traffic monitoring. It was shown that by using different techniques it is possible to classify traffic on links up to 1Gb/s. It looks like the classification of only TCP flows is more efficient - TCP uses session mechanism which allows classification based only on a few packets rather than on the whole flow [4]. Also there are papers devoted to use of SVM in SDN networks. In [9] STIC mechanism was proposed. It is used for internet traffic classification and identification, it classifies 28 different applications. STIC works between the SDN control and forwarding plane. VLAN tagging is used to complete the implementation of different application traffic diversion. Combination of the deep packet inspection and machine learning for application-layer classification was presented in [10]. Authors noticed an increase in classification speed when using more classifiers. Unfortunately, the research also showed a decrease in the performance of the controller on which solution was running.

As depicted there are works regarding use of ML in computer networks. In this paper we present novel look at this topic. We use ML to determine transmission quality rather than to classify type of the traffic. Transmission parameters are measured, preferably at the time that can be considered as real. Based on measured parameters classifier determines transmission quality. It is very difficult to define meaningful compartments of each parameter and then define correlation between them which later will give use information regarding transmission quality. As transmission quality should rely on customer experience classifier should be trained with QoE data. Combination of ML and QoE gives very interesting results. The goal of this paper is to present the novel approach to transmission quality determination, quality which later can be used to provide certain level of the service to the customer. This is the fundamental element of Pay&Require technique.

A. Pay&Require

Pay&Require was proposed as decentralized technique in which customer pays for the transmission quality which is assured. It seems that decentralization is a good direction raising the level of the network security - there is no one central controller whose failure prevents the service from being provided across the entire network. Main assumption of the Pay&Require is that data transmission between customers can be realized through different paths. In the Fig. 1 example of computer network was presented. This network contains four routers (R1-R4), four links between routers (L1-L4) and three customers (C1-C3) connected to the routers - two customers (C1 and C2) connected to R1 and one customer (C3) connected to R3. Transmission quality through link L1-L4 was graded in scale 1-5 - examples of the grades are depicted. Pay&Require allows paths differentiation - different paths from source to destination can be defined based on customer transmission quality expectations. Let's assume that C1 expects transmission quality 5 and C2 accepts transmission

Authorized licensed use limited to: Politechnika Krakowska. Downloaded on September 01,2020 at 09:55:01 UTC from IEEE Xplore. Restrictions apply.

quality 3 on the path to the target which, in this case, is C3. Overall transmission path quality is determined by the lowest transmission quality of the link being part of the path. In presented example there are two possible transmission paths between R1 and R3 i.e. P1={R1-L1-R2-L2-R3} and P2={R1-L3-R4-L4-R3}. P1 has overall transmission quality graded as 5 and P2 has a grade of 3. Thanks to use of Pay&Require it is possible to use path which transmission quality meets customer expectations. In presented example C1 will transmit data to C3 through P1 - overall transmission quality of P1 is 5, C1 expects transmission quality 5. Transmission between C2 and C3 will be done along the path P2 - as per C2 transmission quality expectations. The correct transmission path, that meets customer expectations, is chosen by Pay&Require.

In Pay&Require concept physical plane and control plane were separated but also market plane was defined. The Pay&Require concept is a different approach to computer networks - control plane was decentralized thanks to multi-agent system. Also novelty is market plane separation - all payment mechanisms are outside main network system. This was also achieved by use of multi-agent system in market plane. Market plane makes necessary negotiations with customer and provides information regarding customer and expected transmission quality to control plane. Market plane is flexible, we can define various purchase methods. From very simple to more complex like auctions which allows dynamic pricing of the quality in real time. Pay&Require uses several types of agents:

1) Monitoring
   This type of agent is responsible for transmission quality monitoring. There are one or more instances of this agent in the whole network depending on system configuration. Monitoring agent operates on network device. It is responsible for monitoring transmission parameters within different links. In previous works this agent was acting on the basis of transmission quality parameter (delay, bandwidth, jitter, packet loss ratio) value ranges. Based on predefined ranges agent was able to determine transmission quality in 1–5 grade scale. Parameter ranges were not reliable and because of that in this paper we proposed novel approach which uses ML.
2) Route reconfiguration
   This type of agent is responsible for the reconfiguration process. It cooperates with the monitoring agent. When monitoring agent has determined that the transmission quality parameters are differ than expected values it informs route reconfiguration agent that reconfiguration is necessary. The route reconfiguration agent performs necessary action - assigning new paths. New path must meet customer expectations in the field of transmission quality. After new paths determination, the agent sends the new configuration to the dependant device. Monitoring agent and route reconfiguration agent are implemented as a single agent with both functionalities. Fig. 2 depicts the agent operation with the details of the communication between agents. Firstly agent exchange information regarding networks directly connected to the network device. Subsequently, agents send the information regarding networks they have learned to another agents. The exchange of information ends when all agents have information about the network topology. Monitoring agents have information about the transmission quality required by each customer. The transmission path which will be used by the clients is then selected. Agents configure the network devices based on the received information. The monitoring agent periodically verifies the parameters of the device links. Parameters such as bandwidth, delay, jitter and packet loss ratio are determined. When the parameters are not as per customer expectations the routes are reconfigured.
3) Trader
   This type of agent is responsible for the transmission quality trade. In the simplest case, the user pays a certain amount for assured transmission quality. Various methods of market purchasing and negotiations can be implemented.

In Pay&Require customer pays for transmission parameters that are guaranteed by the system. If parameters are not as customer expects then agents decide if network reconfiguration is necessary. Such reconfiguration means that transmission quality of each route from source to destination must be measured and based on that routing tables are build and applied. After reconfiguration customer transmission should work with expected quality.

Transmission quality was defined in 1-5 scale within Pay&Require. It is very useful and easy for customers. In a very simple way customer knows for what has he paid. Definition of transmission parameters, which afterwards will be recalculated to mentioned scale, is very difficult. There are four base transmission parameters which affect the overall transmission - bandwidth, delay, jitter and packet loss ratio. Choosing a combination of parameter values and converting them into transmission quality scale was always problematic and intricate. In this work we proposed method for transmission quality determination with use of QoE (Quality of Experience) and machine learning (ML).

Generic model of Pay&Require is presented in Fig. 3. In this model, three planes are defined. Plane 1 is a physical plane in which the network devices are operating. Plane 2 is responsible for control. In this plane software agents are implemented for monitoring the transmission quality and, if necessary, for the reconfiguration of the network architecture. Plane 3 is responsible for the interactions with the customer–end user. Assured transmission parameters (quality) purchase techniques are implemented in this plane. In this paper we focus on the modification and extension of the Plane 2 in that model. We implement the software agents responsible for measurement of the transmission parameters. We also implemented the Machine Learning (ML) algorithms, which are used for grading on a certain scale the transmission qual-

Authorized licensed use limited to: Politechnika Krakowska. Downloaded on September 01,2020 at 09:55:01 UTC from IEEE Xplore. Restrictions apply.

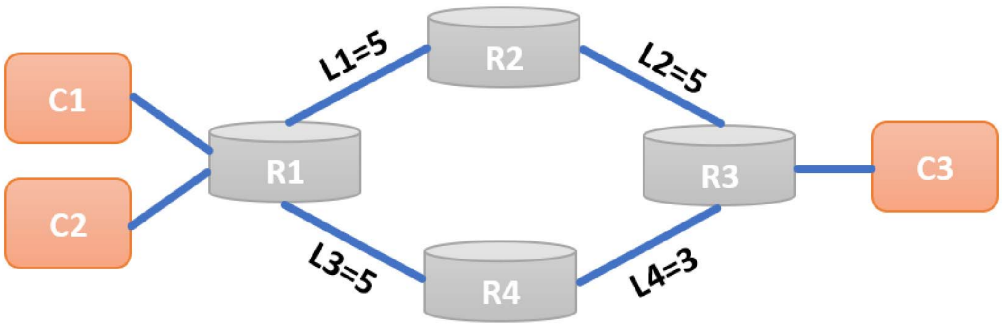


Fig. 1. An example computer network for the Pay&Require technique

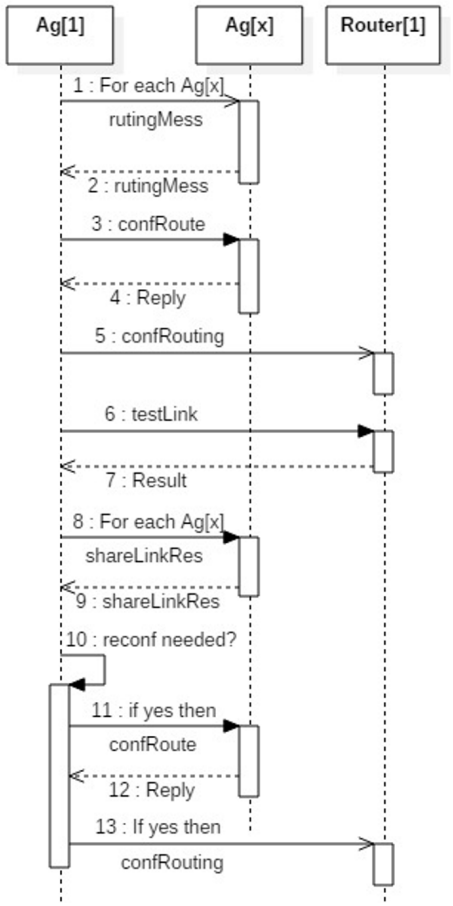


Fig. 2. Agent operation

ity based on transmission parameters. Assessed transmission quality is compared with customer expectations, if it doesn't meet them then agents start reconfiguration process.

III. Developed Network Model Essentials

A. General concept

- This paper presents a new method to determine transmission quality in grade of 1-5 based on bandwidth, delay, jitter and packet loss ratio values.
- The goal is being achieved by use of classifier trained with QoE data. Thanks to QoE it is possible to provide service on the quality level expected by customer.
- Quality level is used by Pay&Require which differentiates the transmission routes based on market oriented techniques. Customer pays for the transmission quality which is guaranteed.

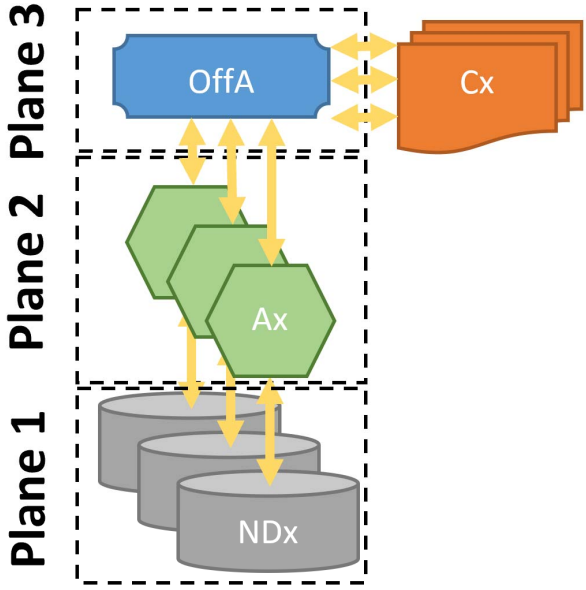


Fig. 3. Pay&Require model

B. Dataset

For research purposes, QoE data was used - thanks to special system prepared for this purpose. First of all it was necessary to measure reliable transmission parameters - bandwidth, delay, jitter and packet loss ratio. To measure mentioned parameters well known tools were used, i.e. iperf, ping, file transmission. After checking reliability of measured parameters it was important to find good network traffic generator which will affect transmission quality. It was decided to use Trex which is open source, low cost, stateful and

Authorized licensed use limited to: Politechnika Krakowska. Downloaded on September 01,2020 at 09:55:01 UTC from IEEE Xplore. Restrictions apply.

TABLE I
DETAILS OF THE DATASET

Number of people tested	4	
Important parameters	Total - 5	
	Input: - Delay - Jitter - Bandwidth - Packet Loss	Output: - Quality
Number of classes	5	
Samples	Class 1: 15 Class 2: 20 Class 3: 26 Class 4: 20 Class 5: 19 Total - 100	

stateless traffic generator. Trex generates traffic specific for Layer 4 - 7 of OSI reference model. It can scale up to 200 Gb/s. Trex contains a lot of examples of real captured traffic. Thanks to that it was possible to simulate real network traffic which greatly affected transmission parameters. The next step was creation of environment in which test user will be able to evaluate transmission quality.

As general decision was to use QoE it was necessary to prepare user-friendly solution. Such solution should contain well known services like video streaming and web page - easy to evaluate by users who are not IT related. In order to do that 100 different scenarios were recorded. One scenario means one combination of measured traffic parameters which have influence on user perception. Based on different parameters, different streaming time and web page loading time was achieved. These different scenarios were recorded and showed to the users. For the purpose of this article four users were asked for their experience. There were two possibilities to measure experience:

- Show each scenario recording and ask for user feeling - transmission quality grade from 1 to 5 (1 - worse quality, 5 - best quality),
- First show reference scenario gained with middle quality and then show test sample and ask user for his feeling in scale -2 to 2 (-2 much worse quality, 2 - much better quality.

In this article first option was chosen. User should evaluate what he sees without any reference. After obtaining results from all users check was done - if there are no huge differences in grades given by different users. As there were no such differences the final result was calculated as average of the grades given by all users and rounded. Details of the dataset used in ML process were presented in table I

C. Methods

There were different methods applied in order to select the best one. Most important parameters for best solution determination were: (a) evaluation of the errors in different classification methods (confusion matrices), and (b) overall accuracy ratio (SEN - sensitivity), which should be the highest.

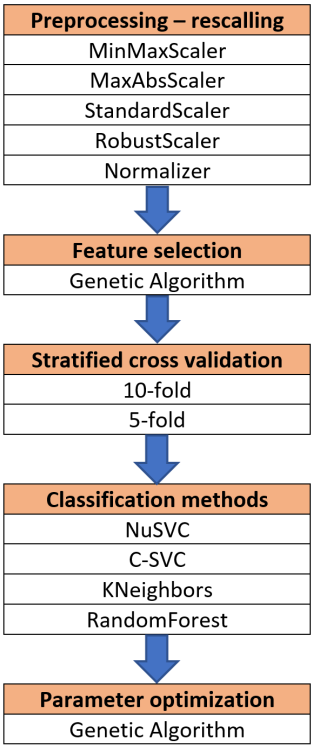


Fig. 4. Organization of the ML system (process).

As a result the best algorithms were chosen - lowest errors, higher overall accuracy. Whole process was divided into five steps, as presented in the Fig. 4.

1) Preprocessing: Different rescalling algorithms were tested in order to achieve valuable preprocessing. Generally rescalling is aimed to get data in the specific range. All possibilities were tested in combination with two different stratified cross validations (CV).

2) Feature selection: In this paper GA (genetic algorithm) feature selection was used to choose the most important features from the whole set. Feature selection is very valuable technique because in some of the cases it is possible to reduce necessary data by eliminating insignificant features. For this purpose Genetic Algorithm was used. Genes of the population of individuals are represented by single attributes/parameters of the transmission given as input for classifiers. Such genes use values of 0 or 1 in order to determine if feature should be applied (1) or rejected (0).

3) Stratified cross validation: Two types of stratified n-fold cross validation were used. Both were tested in all possible combinations of preprocessing, genetic methods and classifiers. We use whole dataset collected based on QoE (100 samples). On mentioned dataset we applied stratified n-fold cross validation. As a result, in the first case 10 combinations (10-fold CV) of testing and training data sets were created, in the second case 5 combinations (5-fold CV). Results are presented only for the test sets of data.

Authorized licensed use limited to: Politechnika Krakowska. Downloaded on September 01,2020 at 09:55:01 UTC from IEEE Xplore. Restrictions apply.

4) Genetic methods: Four types of classifiers were tested: (a) Nu-Support Vector Classifier (Nu-SVC) [28], [30], (b) C-Support Vector Classifier (C-SVC) [28], [30], (c) K-Nearest Neighbors (kNN) algorithm [12], (d) Random Forest Classifier [29]. Classifiers were tested with different parameters presented in table II.

5) Parameter optimization: Genetic algorithm was used to optimize parameters of the classifiers.

D. Evaluation criteria

In order to evaluate the performance, different metrics can be used, i.e. Precision, Specificity, Accuracy, F1 Score, Sensitivity, Matthews Correlation Coefficient etc. In this study we used, appropriate for multi-class problem and WTA rule (winner takes all), overall accuracy (O_ACC) = sensitivity (SEN) [24]–[26]. Overall accuracy, used in this research, was parameter obtained with use of sklearn accuracy score function. Sensitivity (overall accuracy) was calculated based on confusion matrix, as follows:

$$SEN = \frac{\sum TP}{\sum TP + \sum FN} \tag{1}$$

where:

- TP - number of True Positives,
- FN - number of False Negatives.

IV. Experimental Analysis

The proposed solution was implemented in Python with sklearn library. The calculations were performed on different machines, their parameters are not important as in this paper we are not focusing on time but rather on performance. All results were presented for the test set of data. Each classifier was tested with a set of five different rescalling algorithms, two types of stratified cross validation, two different sizes of populations and iterations (100 and 1000). This gave us 80 different combinations which were tested. In this paper three best results for each classifier were presented. In table III the best results for kNN were presented. The best accuracy was always 84%, it was achieved with all features, for 10-fold stratified cross validation. The same accuracy was achieved for MaxAbsScaler, MinMaxScaler and MaxAbsScaler rescalling algorithm. Features are delay, jitter, bandwidth and packet loss ratio, "1" means that feature was taken into consideration, "0" means that feature was eliminated. Results for Nu-SVC classifier were presented in table IV. In case of Nu-SVC the best accuracy was 87%. This result was achieved for 10-fold stratified cross validation, with MinMaxScaler rescalling method and using only three features. Packet loss ratio was eliminated by GA as insignificant feature. This is very interesting result, first of all because of achieved accuracy but secondly because of feature elimination. In table V results for C-SVC were depicted. Best accuracy for C-SVC classifier was achieved for 10-fold stratified cross validation and StandardScaler rescalling method, and it was 83%. In this case very interesting was feature elimination - the best result with use of C-SVC was achieved with only two features: jitter and bandwidth. Such fact should be tested in real system but it looks like there is no possibility to use only these two parameters to determine transmission quality. Results for Random Forest were presented in table VI. Best accuracy was 87%. Such accuracy was achieved for 10-fold stratified cross validation and MaxAbsScaler rescalling method. All four features were used in this case.

For the best result of each classifier confusion matrices were depicted in the Fig. 5, Fig. 6, Fig. 7 and Fig. 8. As problem was more complex than the binary problems it was necessary to prepare general matrix and then extend information for each class (transmission quality rate). True Positives (TP), True Negatives (TN), False Positives (FP), False Negatives (FN) and Sensitivity (SEN) were presented in the tables. Based on SEN it is possible to see for which classes classifier works better and when it works worse.

V. Discussion

A. Hypothesis

Concept presented in this paper is the novel view in transmission quality determination. Normally transmission quality is described by various parameters like bandwidth, delay, jitter and packet loss ratio. Such parameters are very useful for network equipment. They can be used in case of applying QoS or finding problems in the network. Pay&Require concept was presented as novel approach to QoS in computer networks. Pay&Require is combination of multi-agent system and routing based on policies. Customer traffic is differentiated by use of various transmission paths. In this concept customer pays for transmission quality which is guaranteed. It might be difficult for customer to understand how different transmission parameters influence their transmission - e.g. browsing the web pages.

In Pay&Require transmission quality is represented by grade in a scale 1-5. It is more understandable for the customer. Even if it is necessary there might be samples of the video streaming or web page loading to show each quality level. Transmission quality, which until now was defined by a few parameters, is now specified by grade. Normally in such cases conversion tables should be used. Such tables does not seem to be good solution because it is difficult to determine how transmission quality parameters are correlated. Interesting approach is use of ML to get transmission quality grade based on a given (measured) parameters. In this paper such approach was presented. First of all it was necessary to get samples which can be used in ML process. 100 different samples were created and showed to the users. Samples were defined by transmission quality parameters but users seen only web page and video. Based on their experience users rated the quality.

Data obtained from the users was used in ML process. Four classifiers with different parameters were used. Results were presented in this paper. The same, best result, was achieved for:

- Nu-SVC classifier with MinMaxScaler rescalling, 10-times stratified cross validation, kernel rbf, Nu 0,167114,

Authorized licensed use limited to: Politechnika Krakowska. Downloaded on September 01,2020 at 09:55:01 UTC from IEEE Xplore. Restrictions apply.

TABLE II
PARAMETERS USED IN CLASSIFIERS

Classifier	Parameter 1	Parameter 2	Parameter 3	Parameter 4
Nu-SVC	kernel=(linear, rbf, poly, sigmoid)	nu=(0,001-0,5)	degree=(1-5)	gamma=(0,001-2)
C-SVC	kernel=(linear, rbf, poly, sigmoid)	coef0=(0-0.5)	degree=(1-6)	gamma=(0,001-2)
kNN	neighbors=(1-7)	weights=(uniform, distance)	algorithm=(ball tree, kd tree, brute, auto)	leaf size=(15-45)
Random Forest	n estimators=(50-200)	max depth=(1-20)	random state=(0-20)	max samples=(1-80)

TABLE III
THREE BEST RESULTS FOR KNN ALGORITHM

Validation	Rescalling	Neigbors	Weights	Algorithm	Leaf size	Delay	Jitter	Bandwidth	Packet loss	O_ACC = SEN
10	MaxAbsScaler	3	uniform	brute	15	1	1	1	1	84 %
10	MinMaxScaler	3	uniform	brute	34	1	1	1	1	84 %
10	MaxAbsScaler	3	uniform	auto	45	1	1	1	1	84 %

TABLE IV
THREE BEST RESULTS FOR NU-SVC CLASSIFIER

Validation	Rescalling	Kernel	Nu	Degree	Gamma	Delay	Jitter	Bandwidth	Packet loss	O_ACC = SEN
10	MinMaxScaler	rbf	0,167114	4	0,016615	1	1	1	0	87 %
10	StandardScaler	poly	0,219859	1	0,74656	1	1	1	1	85 %
5	MinMaxScaler	poly	0,068178	1	1,197787	1	1	1	0	85 %

TABLE V
THREE BEST RESULTS FOR C-SVC CLASSIFIER

Validation	Rescalling	Kernel	Coef0	Degree	Gamma	Delay	Jitter	Bandwidth	Packet loss	O_ACC = SEN
10	StandardScaler	poly	0,11541	6	1,910356	0	1	1	0	83 %
5	Normalizer	poly	0,19011	6	1,730518	0	1	1	1	82 %
10	Normalizer	poly	0,374051	3	2,37606	1	1	1	1	82 %

TABLE VI
THREE BEST RESULTS FOR RANDOM FOREST CLASSIFIER

Validation	Rescalling	Estimators	Max depth	Random state	Max samples	Delay	Jitter	Bandwidth	Packet loss	O_ACC = SEN
10	MaxAbsScaler	149	4	0	79	1	1	1	1	87 %
10	MinMaxScaler	164	5	14	72	1	1	1	1	86 %
5	MaxAbsScaler	78	17	16	18	1	1	1	1	84 %

Actual \ Predicted	1	2	3	4	5
1	**15**	0	0	0	0
2	3	**15**	2	0	0
3	0	0	**25**	1	0
4	0	0	4	**12**	4
5	0	0	0	2	**17**

	1			2			3			4			5		
1	**15**	0	100,00%	**15**	5	75,00%	**25**	1	96,15%	**12**	8	60,00%	**17**	2	89,47%
			0,00%			25,00%			3,85%			40,00%			10,53%
2	3	**82**	96,47%	0	**80**	100,00%	6	**68**	91,89%	3	**77**	96,25%	4	**77**	95,06%
			3,53%			0,00%			8,11%			3,75%			4,94%
3	83,33%	100,00%	**100,00%**	100,00%	94,12%	**75,00%**	80,65%	98,55%	**96,15%**	80,00%	90,59%	**60,00%**	80,95%	97,47%	**89,47%**
	16,67%	0,00%	0,00%	0,00%	5,88%	25,00%	19,35%	1,45%	3,85%	20,00%	9,41%	40,00%	19,05%	2,53%	10,53%

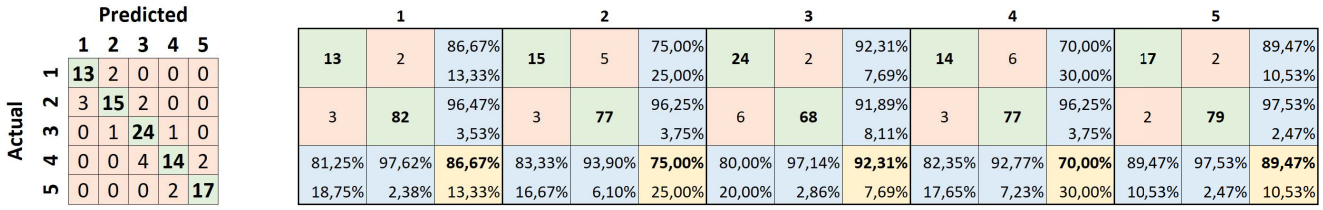
Fig. 5. Confusion matrix for the best kNN algorithm result (SEN = 84%). In yellow cells, SEN was calculated for individual classes.

Actual \ Predicted	1	2	3	4	5
1	**12**	3	0	0	0
2	1	**19**	0	0	0
3	0	1	**23**	2	0
4	0	0	1	**18**	1
5	0	0	0	4	**15**

	1			2			3			4			5		
1	**12**	3	80,00%	**19**	1	95,00%	**23**	3	88,46%	**18**	2	90,00%	**15**	4	78,95%
			20,00%			5,00%			11,54%			10,00%			21,05%
2	1	**84**	98,82%	4	**76**	95,00%	1	**73**	98,65%	6	**74**	92,50%	1	**80**	98,77%
			1,18%			5,00%			1,35%			7,50%			1,23%
3	92,31%	96,55%	**80,00%**	82,61%	98,70%	**95,00%**	95,83%	96,05%	**88,46%**	75,00%	97,37%	**90,00%**	93,75%	95,24%	**78,95%**
	7,69%	3,45%	20,00%	17,39%	1,30%	5,00%	4,17%	3,95%	11,54%	25,00%	2,63%	10,00%	6,25%	4,76%	21,05%

Fig. 6. Confusion matrix for the best Nu-SVC classifier result (SEN = 87%). In yellow cells, SEN was calculated for individual classes.

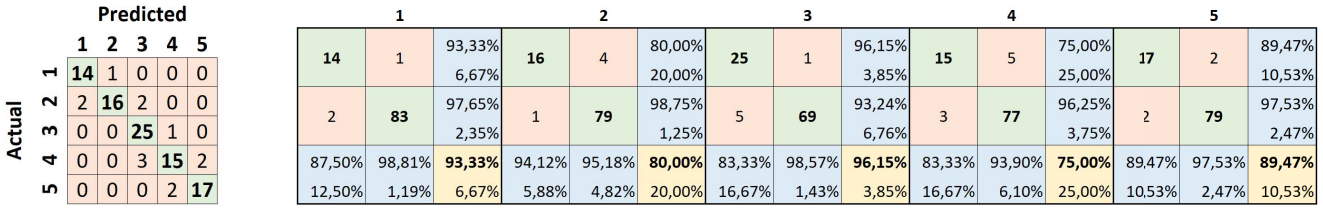
Authorized licensed use limited to: Politechnika Krakowska. Downloaded on September 01,2020 at 09:55:01 UTC from IEEE Xplore. Restrictions apply.



Actual \ Predicted	1	2	3	4	5
1	**13**	2	0	0	0
2	3	**15**	2	0	0
3	0	1	**24**	1	0
4	0	0	4	**14**	2
5	0	0	0	2	**17**

	1			2			3			4			5		
1	**13**	2	86,67%	**15**	5	75,00%	**24**	2	92,31%	**14**	6	70,00%	**17**	2	89,47%
			13,33%			25,00%			7,69%			30,00%			10,53%
2	3	**82**	96,47%	3	**77**	96,25%	6	**68**	91,89%	3	**77**	96,25%	2	**79**	97,53%
			3,53%			3,75%			8,11%			3,75%			2,47%
3	81,25%	97,62%	**86,67%**	83,33%	93,90%	**75,00%**	80,00%	97,14%	**92,31%**	82,35%	92,77%	**70,00%**	89,47%	97,53%	**89,47%**
4	18,75%	2,38%	13,33%	16,67%	6,10%	25,00%	20,00%	2,86%	7,69%	17,65%	7,23%	30,00%	10,53%	2,47%	10,53%

Fig. 7. Confusion matrix for the best C-SVC classifier result (SEN = 83%). In yellow cells, SEN was calculated for individual classes.



Actual \ Predicted	1	2	3	4	5
1	**14**	1	0	0	0
2	2	**16**	2	0	0
3	0	0	**25**	1	0
4	0	0	3	**15**	2
5	0	0	0	2	**17**

	1			2			3			4			5		
1	**14**	1	93,33%	**16**	4	80,00%	**25**	1	96,15%	**15**	5	75,00%	**17**	2	89,47%
			6,67%			20,00%			3,85%			25,00%			10,53%
2	2	**83**	97,65%	1	**79**	98,75%	5	**69**	93,24%	3	**77**	96,25%	2	**79**	97,53%
			2,35%			1,25%			6,76%			3,75%			2,47%
3	87,50%	98,81%	**93,33%**	94,12%	95,18%	**80,00%**	83,33%	98,57%	**96,15%**	83,33%	93,90%	**75,00%**	89,47%	97,53%	**89,47%**
4	12,50%	1,19%	6,67%	5,88%	4,82%	20,00%	16,67%	1,43%	3,85%	16,67%	6,10%	25,00%	10,53%	2,47%	10,53%

Fig. 8. Confusion matrix for the best Random Forest classifier result (SEN = 87%). In yellow cells, SEN was calculated for individual classes.

degree 4, Gamma 0,016615. One feature was eliminated - packet loss ratio.

- Random Forest classifier with MaxAbsScaler rescalling, 10-times stratified cross validation, 149 estimators, max depth 4, random state 0, max samples 79. All features were used.

87% seems to be a good result at this stage of research. As presented problem is not binary, standard accuracy could not be used. To compare results it was necessary to create confusion matrix for each class (transmission quality grade). In the matrix it is easy to see where classifier worked fine and where worse results were achieved. The best classifier was Nu-SVC because, as per confusion matrix, the worst class was classified with SEN = 79%. kNN had SEN even up to 100% but it also has the highest SEN spread. In some classes SEN was very high but in other ones it was very low. Very interesting was the fact that highest SEN was achieved for classifier using only three features, one feature was eliminated as not relevant. Conclusions are that presented approach is very interesting and might be very useful in the systems where we have few different parameters and we want to move to easy grading. Thanks to ML it is possible to have very flexible and useful system.

VI. Conclusions

In this paper we presented (1) data set obtained thanks to users experience. Data can be used as a source of correlation between transmission quality parameters and grade. (2) novel approach which allows conversion from transmission quality parameters to defined grade scale with use of ML. (3) application of genetic methods coupled with feature selection and cross validation optimized for transmission quality classification. Our solution uses ML which is very good for this purpose. Our research showed that ML can be used and might be very useful in case of network in which Pay&Require concept is used. It also seems that such ML classification can be used in other software defined networks. Advantages of presented approach are that it is flexible and valuable if used in computer networks. It simplifies transmission quality assessment and makes evaluation by the customers easier. It appears that use of ML can provide a more reliable transmission quality converter.

Disadvantage is that it is necessary to have enough reliable samples which mean that grades given by test users cannot be accidental. Users giving samples must be credible and the whole system is based on their experience. Another disadvantage is the number of examples given to the users to evaluate. In computer networks small differences in transmission quality parameters did not affect examples showed to the test users. So test examples should be created very carefully and should be checked by a person with relevant experience before examining the test users.

In this paper proposed ML solution was limited only to use in Pay&Require which is novel approach to transmission quality assurance in computer networks. Also samples were limited - 100 samples and 4 users tested. In the future works there should be more test users. In this paper we focused on four classifiers (1) Nu-SVC, (2) C-SVC, (3) kNN and (4) Random Forest. In future works other classifiers and techniques should be tested to get better ACC and SEN. Currently used classifiers should be tested with other values of the parameters. 87% of classification accuracy is quite good result but in future works it seems to be possible to get better results which might be very useful in a real system. Generally proposed method looks very promising, especially at this stage of research. It should be deeply tested in order to obtain better classification.

Acknowledgement

Joanna Kolodziej's work was supported in part by the European Commission, under Grant Agreement no. 833456.

Authorized licensed use limited to: Politechnika Krakowska. Downloaded on September 01,2020 at 09:55:01 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] Żelasko, D.; Cetnarowicz, K.; Wajda, K.; Koźlak, J.: Pay&Require as concept of variable cost routing in dynamically reconfigured networks, Technical Transactions, 2016.

[2] Żelasko, D.: Ensuring the QoS in computer networks through the use of the Pay&Require multi-agent system and electronic auctions, Technical Transactions, 2018.

[3] Software-Defined Networking (SDN): Layers and Architecture Terminology (RFC7426).

[4] Este, A.; Gringoli, F.; Salgarelli, L.: Support Vector Machines for TCP traffic classification, Computer Networks, Volume 53, Issue 14, pages 2476-2490, 2009.

[5] Li, Z.; Yuan, R.; Guan, X.: Accurate Classification of the Internet Traffic Based on the SVM Method, IEEE International Conference on Communications, 2007.

[6] Yuan, R.; Li, Z.; Guan X.; Xu, L.: An SVM-based machine learning method for accurate internet traffic classification, Information Systems Frontiers, volume 12, pages 149–156, 2010.

[7] Este, A.; Gringoli, F.; Salgarelli, L.: On-line SVM traffic classification, 7th International Wireless Communications and Mobile Computing Conference, 2011.

[8] Nguyen, T.T.T.; Armitage, G.: A survey of techniques for internet traffic classification using machine learning, IEEE Communications Surveys Tutorials, volume: 10 , issue: 4 , pages 56-76, 2008.

[9] Liu, C.-C.; Chang, Y.; Tseng, C.-W.; Yang, Y.-T.; Lai, M.-S.; Chou, L.-D.: SVM-based Classification Mechanism and Its Application in SDN Networks, 10th International Conference on Communication Software and Networks (ICCSN), 2018.

[10] Li, Y.; Li, J.: MultiClassifier: A combination of DPI and ML for application-layer classification in SDN, 2nd International Conference on Systems and Informatics (ICSAI 2014), 2014.

[11] Pławiak, P.; Abdar, M.; Acharya, U.R.: Application of New Deep Genetic Cascade Ensemble of SVM Classifiers to Predict the Australian Credit Scoring, Applied Soft Computing, Volume 84, November 2019.

[12] Altman, N.S.: An introduction to kernel and nearest neighbor non parametric regression. Am Stat 46(3):175–185, 1992.

[13] Tadeusiewicz, R.: Neural networks as a tool for modeling of biological systems, BIO-ALGORITHMS AND MED-SYSTEMS, Volume: 11, Issue: 3, Pages: 135-144, 2015.

[14] Back, T; Hammel, U; Schwefel, H.P.: Evolutionary computation: comments on the history and current state, IEEE TransEvol Comput 1(1):3–17, 1997.

[15] Szaleniec, M.; Tadeusiewicz, R.; Witko, M.: How to select an optimal neural model of chemical reactivity?, NEUROCOMPUTING, Volume: 72, Issue: 1-3, Special Issue: SI, Pages: 241-256, 2008.

[16] Cortes, C.; Vapnik, V.: Support-vector networks, Mach Learn20(3):273–297, 1195.

[17] Cruz, R.M.; Sabourin, R.; Cavalcanti, G.D.: Dynamic classifier selection: recent advances and perspectives, Inf Fusion 41(Supplement C):195–216, 2018.

[18] Szaleniec, J.; Wiatr, M.; Szaleniec, M.; Składzień, J.; Tomik, J.; Oleś, K.; Tadeusiewicz, R.: Artificial neural network modelling of the results of tympanoplasty in chronic suppurative otitis media patients, COMPUTERS IN BIOLOGY AND MEDICINE, Volume: 43, Issue: 1, Pages: 16-22, 2013.

[19] Oza, N.C.; Tumer, K.: Classifier ensembles: select real-world applications, Inf Fusion 9(1): 4–20, 2008.

[20] Pławiak, P.; Maziarz, W.: Classification of tea specimens using novel hybrid artificial intelligence methods, Sensors Actuators B Chem 192:117–125, 2014.

[21] Plawiak, P.; Tadeusiewicz, R.: Approximation of phenol concentration using novel hybrid computational intelligence methods, APPLIED MATHEMATICS AND COMPUTER SCIENCE, Volume: 24, Issue: 1, Pages: 165-181, 2014.

[22] Rutkowski, L.: Computational intelligence: methods and techniques, Springer, Berlin, 2008.

[23] Sokolova, M.; Lapalme, G.: A systematic analysis of performance measures for classification tasks, Inf Process Manag 45(4):427–437, 2019.

[24] Pławiak, P.; Acharya, U.R.: Novel deep genetic ensemble of classifiers for arrhythmia detection using ECG signals, Neural Computing and Applications, pages 1-25, 2018.

[25] Pławiak, P.: Novel genetic ensembles of classifiers applied to myocardium dysfunction recognition based on ECG signals, Swarm and Evolutionary Computation, Volume 39, Pages 192-208, 2018.

[26] Pławiak, P.: Novel methodology of cardiac health recognition based on ECG signals and evolutionary-neural system, Expert Systems with Applications, 92C(2018): 334-349, 2018.

[27] Kuncheva, L. I.: Combining Pattern Classifiers: Methods and Algorithms, Wiley-Interscience, 2004.

[28] Chang, C. C.; Lin, C. J: LIBSVM: A library for support vector machines, ACM Trans. Intell. Syst. Technol. (TIST) 2 (2011) 27:1–27:27, Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[29] Breiman, L.: Random Forests, Machine Learning, 45(1), 5-32, 2001.

[30] Huang, C. L.; Chen, M. C.; Wang, C. J.: Credit scoring with a data mining approach based on support vector machines, Expert Syst. Appl. 33 (4) (2007) 847–856

Copy protected with Online-PDF-No-Copy.com

Authorized licensed use limited to: Politechnika Krakowska. Downloaded on September 01,2020 at 09:55:01 UTC from IEEE Xplore. Restrictions apply.